# ITERATION-COMPLEXITY OF AN INNER ACCELERATED INEXACT PROXIMAL AUGMENTED LAGRANGIAN METHOD BASED ON THE CLASSICAL LAGRANGIAN FUNCTION

WEIWEI KONG[1], JEFFERSON G. MELO[2], AND RENATO D.C. MONTEIRO[3]

**Abstract.** This paper establishes the iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian (IAIPAL) method for solving linearly-constrained smooth nonconvex composite optimization problems that is based on the classical augmented Lagrangian (AL) function. More specifically, each IAIPAL iteration consists of inexactly solving a proximal AL subproblem by an accelerated composite gradient (ACG) method followed by a classical Lagrange multiplier update. Under the assumption that the domain of the composite function is bounded and the problem has a Slater point, it is shown that IAIPAL generates an approximate stationary solution in $\mathcal{O}(\varepsilon^{-5/2}\log^2\varepsilon^{-1})$ ACG iterations where $\varepsilon > 0$ is a tolerance for both stationarity and feasibility. Moreover, the above bound is derived without assuming that the initial point is feasible. Finally, numerical results are presented to demonstrate the strong practical performance of IAIPAL.

**Key words.** inexact proximal augmented Lagrangian method, linearly constrained smooth nonconvex composite programs, inner accelerated first-order methods, iteration complexity.

**AMS subject classifications.** 47J22, 49M27, 90C25, 90C26, 90C30, 90C60, 65K10.

## 1. Introduction.

This paper presents an inner accelerated inexact proximal augmented Lagrangian (IAIPAL) method for solving the linearly-constrained smooth nonconvex composite optimization (NCO) problem

$$\phi^* := \min\{\phi(z) := f(z) + h(z) : Az = b\}, \tag{1.1}$$

where $A : \Re^n \to \Re^l$ is a linear operator, $b \in \Re^l$, $h : \Re^n \to (-\infty, \infty]$ is a closed proper convex function which is $M_h$-Lipschitz continuous on its domain, and $f$ is a real-valued differentiable nonconvex function such that, for some scalars $L_f \geq m_f > 0$, $f$ is $m_f$-weakly convex on the domain, $\operatorname{dom} h$, of $h$ (i.e., satisfies (2.2) below) and its gradient is $L_f$–Lipschitz. For a given tolerance pair $(\hat\rho, \hat\eta) \in \Re^2_{++}$, its goal is to find a triple $(\hat z, \hat p, \hat w)$ satisfying

$$\hat w \in \nabla f(\hat z) + \partial h(\hat z) + A^*\hat p, \quad \|\hat w\| \leq \hat\rho, \quad \|A\hat z - b\| \leq \hat\eta. \tag{1.2}$$

More specifically, IAIPAL is based on the augmented Lagrangian (AL) function $\mathcal{L}_c(z; p)$ defined as

$$\mathcal{L}_c(z; p) := f(z) + h(z) + \langle p, Az - b\rangle + \frac{c}{2}\|Az - b\|^2, \tag{1.3}$$

which has been thoroughly studied in the literature (see for example [3, 5, 22, 28, 38]). Roughly speaking, for a fixed stepsize $\lambda > 0$, a scalar $\alpha > 0$, and initial points

[1]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37830. (email: wwkong92@gmail.com). This author has been supported by (i) the US Department of Energy (DOE) and UT-Battelle, LLC, under contract DE-AC05-00OR22725 and (ii) the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

[2]Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (email: jefferson@ufg.br). This author was partially supported by CNPq grant 312559/2019-4 and FAPEG/GO.

[3]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). This author was partially supported by ONR Grant N00014-18-1-2077 and AFORS Grant FA9550-22-1-0088.

$z_0 \in \operatorname{dom} h$ and $p_0 = 0$, IAIPAL repeatedly performs the following iteration: given $(z_{k-1}, p_{k-1}) \in \operatorname{dom} h \times \Re^l$, it computes $(z_k, p_k)$ as

$$z_k \approx \operatorname*{argmin}_z \left\{ \lambda \mathcal{L}_c(z, p_{k-1}) + \frac{1}{2}\|z - z_{k-1}\|^2 \right\} \quad , \tag{1.4}$$

$$p_k = p_{k-1} + \begin{cases} c(Az_k - b), & k \equiv 1 \bmod \lceil \alpha c \rceil, \\ 0, & \text{otherwise}, \end{cases} \tag{1.5}$$

where $z_k$ in (1.4) is a suitable approximate solution of the underlying prox-AL subproblem (1.4). IAIPAL sets $\lambda = 1/(2m_f)$ which, due to the fact that $f$ is $m_f$-weakly convex, guarantees that the objective function of (1.4) is strongly convex. Moreover, it computes $z_k$ by approximately solving subproblem (1.4) by a strongly convex version of FISTA, which is a well-known accelerated composite gradient (ACG) variant for solving convex composite optimization problems (see for example [4, 31, 33]). The latter point is then used to construct a triple $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ and IAIPAL stops if it satisfies (1.2). Otherwise, an auxiliary novel test is performed to decide whether: i) $c$ should be left unchanged, or; ii) $c$ is updated as $c \leftarrow \tau c$ with $\tau > 1$ and $(z_k, p_k)$ either changed to $(z_0, p_0)$ (cold restart) or to $(z_k, p_0)$ (warm restart). Finally, $k$ is updated to $k + 1$ and the iteration described above is repeated.

*Related works.* We mainly focus our attention on works dealing with iteration complexities of penalty-based and/or AL-based methods. Furthermore, all of the AL methods below use the multiplier update

$$p_k = (1 - \theta)(p_{k-1} + \chi_k c_k (Az_k - b)) \tag{1.6}$$

for $\theta \in [0, 1)$ and $\chi_k \in [0, 1]$ at every $k \geq 1$. For consistency, the complexities in this review refer to the effort of obtaining an approximate stationary point as in (1.2). Note that even though these complexities are described as bounds on the number of (possibly ACG) iterations, they are also bounds on the total number of $h$-resolvent computations and/or gradient evaluations of $f$.

Iteration complexities of quadratic penalty methods for solving (1.1) under the assumption that $f$ is convex and $h$ is an indicator function of a convex set were first analyzed in [21] and further studied in [2, 32]. Iteration complexities of first-order augmented Lagrangian (AL) methods for solving the aforementioned class of convex problem have been studied in [3, 22, 23, 28, 29, 36, 41]. Proximal quadratic penalty (PQP) methods were first studied in [18] and further developed in [19, 20, 26].

Classical proximal AL (PAL) methods for solving (1.1) under the assumption that $f$ is convex, $\theta = 0$, and $\chi_k = 1$ for every $k$ have first been studied in [39]. Recently, papers [9, 30] studied PAL methods under the assumption that $f$ is (possibly) nonconvex, $\theta \in (0, 1]$, and $\chi_k = 1$ for every $k$. However, as $\theta$ approaches zero, the prox stepsize $\lambda$ of these methods converge to zero which causes the following issues: (i) their derived complexity bounds diverge to infinity (see the second column in Table 1.2 below), which invalidates their analyses for the case of $\theta = 0$; and (ii) deteriorating computational performance.

Papers [24, 40] study non-proximal AL methods under the assumption that $f$ is nonconvex, $\theta = 0$, and $\chi_k = O(c_k^{-1})$ for every $k$. It is worth mentioning that both [24, 40] make a strong assumption about the generated iterates (see condition $\mathcal{F}$ in Table 1.1), which have only been shown to hold when $h$ is the indicator of a polyhedron or a ball. Moreover, [40] considers the more general version of (1.1) where the constraints are (possibly) nonconvex.

We now describe other papers that are tangentially related to ours. Papers [42, 43] present a primal-dual first-order algorithm under the assumption that $h$ is the indicator function of a box (in [43]) or a polyhedron (in [42]). Paper [15] considers a penalty-ADMM method that solves an equivalent reformulation of (1.1). Paper [25] presents an inexact proximal point method applied to the function defined as $\phi(z)$ if $z$ is feasible and $+\infty$ otherwise. It can be viewed as an extension to the nonconvex setting of the proximal point method (PPM) applied to (1.1) (see, for example, [39] for the analysis of inexact versions of PPMs for solving (1.1) in the convex setting). Paper [6] considers a primal-dual proximal point scheme for computing approximate stationary solution to a constrained NCO problem and analyzes its iteration-complexity under different assumptions.

Before closing this review, we present the assumptions of some of the above methods in Table 1.1 and give a summary of these methods in Table 1.2, which compares the best iteration complexities, necessary conditions, and various parameter ranges.

| | |
|---|---|
| $\mathcal{B}$ | Either (i) the quantity $\sup_{x\in\operatorname{dom} h}\|\phi(x)\|$ is finite, (ii) $\operatorname{dom} h$ is bounded, and/or (iii) the feasible set is bounded. |
| $\mathcal{A}$ | If the constraints have an affine component of the form $Ax = b$ then $A$ has full row rank. |
| $\mathcal{F}$ | There exists some $\nu > 0$ such that $\nu\|Ax_k - b\| \le \operatorname{dist}(0, A^*(Ax_k - b) + c_k^{-1}\partial h(x_k))$ for generated iterates $\{x_k\}_{k\ge1}$ and $\{c_k\}_{k\ge1}$. |
| $\mathcal{N}$ | The function $h$ restricted to its domain is $r$-Lipschitz continuous. |
| $\mathcal{SP}$ | There exists $\bar{x} \in \operatorname{int}(\operatorname{dom} h)$ such that $Ax = b$. |

TABLE 1.1

*Abbreviations for common boundedness and regularity conditions. See Lemma A.2 for the result that $\mathcal{N}$ is equivalent to requiring that, for every $x \in \operatorname{dom} h$, there exists $r > 0$ such that that $\partial h(x) \subseteq \mathcal{N}_{\operatorname{dom} h}(x) + \mathcal{B}_r$ where $\mathcal{B}_r = \{x : \|x\| \le r\}$.*

| **Name** | **Best Complexity** | $\lambda_k$ | $\theta$ | $\chi_k$ | **Conditions** |
|---|---|---|---|---|---|
| QP-AIPP [18] | $\mathcal{O}(\varepsilon^{-3})$ | $\Theta(m_f^{-1})$ | - | - | None |
| R-QP-AIPP [19] | $\tilde{\mathcal{O}}(\varepsilon^{-3})$ | $(0,\infty)$ | - | - | $\mathcal{B}$ |
| iPPP [26] | $\tilde{\mathcal{O}}(\varepsilon^{-5/2})$ | $\mathcal{O}(m_f^{-1})$ | - | - | $\mathcal{B},\mathcal{N},\mathcal{SP}$ |
| iALM (2019) [40] | $\tilde{\mathcal{O}}(\varepsilon^{-3})$ | - | 0 | $\mathcal{O}(c_k^{-1})$ | $\mathcal{B},\mathcal{F}$ |
| iALM (2020) [24] | $\tilde{\mathcal{O}}(\varepsilon^{-5/2})$ | - | 0 | $\mathcal{O}(c_k^{-1})$ | $\mathcal{B},\mathcal{F}$ |
| PProx-PDA[4] [9] | $\mathcal{O}(\theta^{-2}\varepsilon^{-4})$ | $\mathcal{O}(\theta L_f^{-1})$ | $(0,1)$ | 1 | $\mathcal{B}$, $\mathcal{A}$ |
| $\theta$-IPAAL[5] [30] | $\tilde{\mathcal{O}}(\theta^{-15/4}\varepsilon^{-5/2})$ | $\Theta(\theta m_f^{-1})$ | $(0,1)$ | 1 | $\mathcal{N},\mathcal{SP}$ |
| **IAIPAL** | $\tilde{\mathcal{O}}(\varepsilon^{-5/2})$ | $\Theta(m_f^{-1})$ | 0 | $\{0,1\}$[6] | $\mathcal{B},\mathcal{N},\mathcal{SP}$ |

TABLE 1.2

*Comparison of relevant penalty and AL-based methods with IAIPAL. For simplicity, we let $\varepsilon = \min\{\hat{\rho},\hat{\eta}\}$, and let $\tilde{\mathcal{O}}(\cdot)$ be the same as $\mathcal{O}(\cdot)$ with all logarithmic dependencies on $\varepsilon$ removed.*

*Contributions.* Under the assumption that the domain of $h$ is bounded, has nonempty interior, and (1.1) has a point $\bar{z} \in \operatorname{int}(\operatorname{dom} h)$ such that $A\bar{z} = b$, it is shown that if $\alpha = \Theta(1)$ then the total ACG iteration complexity of IAIPAL, up to logarithmic terms, is

$$\mathcal{O}\left(\frac{1}{\hat{\rho}^{5/2}} + \frac{1}{\sqrt{\hat{\eta}}\hat{\rho}^2}\right). \tag{1.7}$$

which is equal, up to logarithmic terms, to the ones obtained for the methods in [24, 26, 30] (see Table 1.2). On the other hand, if $\alpha = 1/c$, i.e., a full multiplier update is performed at every step of IAIPAL in view of (1.5), then it is shown that the above complexity becomes $\mathcal{O}(\hat{\rho}^{-3} + \hat{\rho}^{-2}\eta^{-1/2})$. Since each ACG iteration of IAIPAL requires $\mathcal{O}(1)$ resolvent evaluations of $h$ and/or gradient evaluations of $f$, the above complexities also bound the number of $h$-resolvent computations (i.e., evaluations of $(I + \eta\partial h)^{-1}$ for $\eta > 0$) and gradient evaluations of $f$ performed by IAIPAL. It is also worth mentioning that all of the above results hold without assuming that the initial point $z_0 \in \operatorname{dom} h$ is feasible, i.e., $z_0$ also satisfies $Az_0 = b$.

We now emphasize four important theoretical aspects of this paper. First, it establishes (for the *first* time) the iteration-complexity of a PAL method (specifically IAIPAL with $\alpha = 1/c$) for solving (1.1) which makes a full multiplier update (i.e., $(\theta, \chi_k) = (0, 1)$ for every $k$) after solving each prox subproblem, does not assume boundedness of the multiplier sequence $\{p_k\}$, and contains a novel rule for updating the penalty parameter. Second, the proof that the sequence of Lagrange multipliers is bounded does not use potential function arguments (e.g. as in [9, 14, 30]), restrict the size of $\chi_k$ in (1.6) (e.g. as in [24, 40]), and/or limit the number of multiplier updates (e.g. as in [24, 40]). Third, in contrast to the PAL methods of [9, 30], whose iteration-complexities and stepsizes tend to $\infty$ and 0, respectively, as $\theta$ tends to 0 (see the second column in Table 1.2), the complexity and stepsize of IAIPAL do not depend on $\theta$. Fourth, in contrast to the non-proximal AL methods of [24, 40], which make strong assumptions on the generated iterates (see condition $\mathcal{F}$ in Table 1.1), the convergence of IAIPAL only assumes a mild Slater-like condition and Lipschitz continuity of $h$ on its domain.

It is also worth mentioning that the numerical experiments of Section 4, and the conclusions thereof, show that IAIPAL with $\alpha = \Theta(1)$ and $\alpha = \Theta(1/c)$ substantially outperforms other algorithms in the literature for solving (1.1) (or special cases of it) with equal (e.g., [18, 19, 26, 30]) or better (e.g., [42, 43]) iteration complexities.

*Organization of the paper.* Subsection 1.1 provides some basic definitions and notation. Section 2 contains three subsections. The first one presents our main problem of interest and the assumptions made on it. The second one states S-IAIPAL and its main iteration-complexity result. Subsection 2.3 states IAIPAL and establishes its iteration-complexity bound. Section 3 is devoted to the proof of the iteration-complexity result of S-IAIPAL and some related technical results. Section 4 presents some numerical experiments comparing IAIPAL with other benchmarks algorithms

[4] This method generates prox subproblems of the form $\operatorname{argmin}_{x \in X}\{\lambda h(x) + c\|Ax - b\|^2/2 + \|x - x_0\|^2/2\}$ and the analysis of [9] makes the strong assumption that they can be solved exactly for any $x_0$, $c$, and $\lambda$.

[5] It is also shown that conditions $\mathcal{N}$ and $\mathcal{SP}$ can be removed to yield a complexity of $\tilde{\mathcal{O}}(\theta^{-7/2}\varepsilon^{-3})$.

[6] Specifically, $\chi_k = 1$ if $k \equiv 1 \bmod k_\circ$ and $\chi_k = 0$ otherwise.

for solving (1.1). Section 5 contains some concluding remarks. Finally, an appendix section is considered and it is divided into three subsections. Subsection A.1 reviews an ACG method used to solve the S-IAIPAL subproblems. The second subsection contains a basic result of convex analysis, and the last subsection presents a basic lemma associated with a refinement procedure considered in S-IAIPAL.

**1.1. Notation and basic definitions.** This subsection presents notation and basic definitions used in this paper.

Let $\mathbb{N}$ denote the set of positive integers. Let $\Re_+$ and $\Re_{++}$ denote the set of nonnegative and positive real numbers, respectively, and let $\Re^n$ denote the $n$-dimensional Hilbert space with inner product and associated norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. We use $\Re^{l\times n}$ to denote the set of all $l\times n$ matrices and $\mathbb{S}^+_n$ to denote the set of positive semidefinite matrices in $\Re^{n\times n}$. The smallest positive singular value of a nonzero linear operator $Q:\Re^n\to\Re^l$ is denoted by $\sigma^+_Q$. For a given closed convex set $X\subset\Re^n$, its boundary is denoted by $\partial X$ and the distance of a point $x\in\Re^n$ to $X$ is denoted by $\mathrm{dist}_X(x)$. For any $t>0$, we let $\log^+_1(t):=\max\{\log t,1\}$ and $\bar B(0,t):=\{z\in\Re^n:\|z\|\le t\}$.

The domain of a function $h:\Re^n\to(-\infty,\infty]$ is the set $\mathrm{dom}\,h:=\{x\in\Re^n: h(x)<+\infty\}$. Moreover, $h$ is said to be proper if $\mathrm{dom}\,h\neq\emptyset$. The set of all lower semi-continuous proper convex functions defined in $\Re^n$ is denoted by $\overline{\mathrm{Conv}}\,(\Re^n)$. The $\varepsilon$-subdifferential of a proper function $h:\Re^n\to(-\infty,\infty]$ is defined by

$$\partial_\varepsilon h(z):=\{u\in\Re^n: h(z')\ge h(z)+\langle u,z'-z\rangle-\varepsilon,\quad\forall z'\in\Re^n\} \tag{1.8}$$

for every $z\in\Re^n$. The classical subdifferential, denoted by $\partial h(\cdot)$, corresponds to $\partial_0 h(\cdot)$. Recall that, for a given $\varepsilon\ge0$, the $\varepsilon$-normal cone of a closed convex set $C$ at $z\in C$, denoted by $N^\varepsilon_C(z)$, is defined as $N^\varepsilon_C(z):=\{\xi\in\Re^n:\langle\xi,u-z\rangle\le\varepsilon,\quad\forall u\in C\}$. If $\psi$ is a real-valued function which is differentiable at $\bar z\in\Re^n$, then its affine approximation $\ell_\psi(\cdot,\bar z)$ at $\bar z$ is given by

$$\ell_\psi(z;\bar z):=\psi(\bar z)+\langle\nabla\psi(\bar z),z-\bar z\rangle\quad\forall z\in\Re^n. \tag{1.9}$$

**2. The IAIPAL method.** This section is divided into three subsections. The first one discusses the problem of interest and describes the main assumptions made on it. Subsection 2.2 presents S-IAIPAL and its main iteration-complexity result. Subsection 2.3 presents IAIPAL and its overall ACG iteration-complexity result.

**2.1. Problem of interest, assumptions and IAIPAL outline.** This subsection describes the problem of interest, the assumptions made on it, and the type of approximate stationary solution we are interested in computing for it.

The main problem of interest in this paper is (1.1) where $f,h:\Re^n\to(-\infty,\infty]$, $A:\Re^n\to\Re^l$ and $b\in\Re^l$ satisfy the following assumptions:

- **(B1)** $A$ is a nonzero linear operator;
- **(B2)** $h\in\overline{\mathrm{Conv}}\,(\Re^n)$ is $L_h$-Lipschitz continuous on $\mathcal{H}:=\mathrm{dom}\,h$;
- **(B3)** the diameter $D:=\sup\{\|z-z'\|:z,z'\in\mathcal{H}\}$ of $\mathcal{H}$ is finite and there exists $\nabla_f\ge0$ such that $\|\nabla f(z)\|\le\nabla_f$ for every $z\in\mathcal{H}$;
- **(B4)** there exists $\bar z\in\mathrm{int}(\mathcal{H})$ such that $A\bar z=b$;
- **(B5)** $f$ is nonconvex and differentiable on $\Re^n$, and there exist $L_f\ge m_f>0$ such that, for all $z,z'\in\Re^n$,

$$\|\nabla f(z')-\nabla f(z)\|\le L_f\|z'-z\|, \tag{2.1}$$

$$f(z')-\ell_f(z';z)\ge-\frac{m_f}{2}\|z'-z\|^2. \tag{2.2}$$

Some comments about assumptions **(B1)**–**(B5)** are in order. First, it is shown in Lemma A.2 that $\partial_\varepsilon h(z) \subset \bar{B}(0, L_h) + N^\varepsilon_{\mathcal{H}}(z)$ for every $z \in \mathcal{H}$. This inclusion will be used to bound the sequence of Lagrangian multipliers generated by the IAIPAL method. Second, it is well known that (2.1) implies that $|f(z') - \ell_f(z'; z)| \le L_f\|z' - z\|^2/2$ for every $z, z' \in \Re^n$, and hence that (2.2) holds with $m_f = L_f$. However, better iteration-complexity bounds can be derived when a scalar $m_f < L_f$ satisfying (2.2) is available. Third, (2.2) implies that the function $f(\cdot) + m_f\|\cdot\|^2/2$ is convex on $\Re^n$. Moreover, since $f$ is nonconvex on $\Re^n$ in view of **(B5)**, the smallest $m_f$ satisfying (2.2) is positive. Fourth, any function $f$ of the form $h = \tilde{h} + \delta_Z$ where $\tilde{h}$ is a finite everywhere Lipschitz continuous convex function and $Z$ is a compact convex set clearly satisfies condition (B2). Finally, the existence of a scalar $\nabla_f$ as in **(B3)** is actually not an extra assumption since, using (2.1) and the boundedness of $\mathcal{H}$ in **(B3)**, it can be easily seen that for any $y \in \mathcal{H}$, the scalar $\nabla_f = \nabla_{f,y} := \|\nabla f(y)\| + L_f D$ majorizes $\|\nabla f(z)\|$ for any $z \in \mathcal{H}$.

It is well known that, under some mild conditions, if $\bar{z}$ is a local minimum of (1.1), then there exists $\bar{p} \in \Re^l$ such that $(\bar{z}, \bar{p})$ is a stationary solution of (1.1), i.e.,

$$0 \in \nabla f(\bar{z}) + \partial h(\bar{z}) + A^*\bar{p}, \quad A\bar{z} - b = 0. \tag{2.3}$$

The main complexity results of this paper are stated in terms of the following notion of approximate stationary solution which is a natural relaxation of (2.3).

DEFINITION 2.1. *Given a tolerance pair $(\hat{\rho}, \hat{\eta}) \in \Re_{++} \times \Re_{++}$, a triple $(\hat{z}, \hat{p}, \hat{w}) \in \mathcal{H} \times \Re^l \times \Re^n$ is said to be a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1.1) if it satisfies (1.2).*

**2.2. The S-IAIPAL method.** This subsection describes S-IAIPAL, which essentially corresponds to to some group of all consecutive iterations of the general IAIPAL method outlined in Section 1 (see the paragraph containing (1.4)-(1.5)) for which the penalty parameter $c$ stays constant.

Recall from the outline given in the Introduction that S-IAIPAL generates a sequence $\{(z_k, p_k)\}$ according to (1.4) and (1.5) where $\lambda = 1/(2m_f)$. The formal description of S-IAIPAL below requires that, for a pre-specified scalar $\tilde{\sigma} > 0$, the approximate solution $z_k$ of subproblem (1.4), together with some residual pair $(v_k, \varepsilon_k) \in \Re^n \times \Re_{++}$, satisfy

$$v_k \in \partial_{\varepsilon_k}\left(\lambda\mathcal{L}_c(\cdot, p_{k-1}) + \frac{1}{2}\|\cdot - z_{k-1}\|^2\right)(z_k), \qquad \|v_k\|^2 + 2\varepsilon_k \le \tilde{\sigma}^2\|r_k\|^2 \tag{2.4}$$

where

$$r_k := z_{k-1} - z_k + v_k. \tag{2.5}$$

We now make some remarks about the above notion of approximate solution for (1.4). First, even though $\tilde{\sigma}$ is assumed to positive, it is worth noting that if $\tilde{\sigma}$ were equal to zero, then (2.4) would immediately imply that $z_k$ is the exact solution of (1.4). Hence, the aggregated error $\|v_k\|^2 + 2\varepsilon_k$ of the residual pair $(v_k, \varepsilon_k)$ can be thought as an inexactness measure of the approximate solution $z_k$, and the inequality in (2.4) is a relative error condition on it. Second, as will be seen in Proposition 2.2 below, a triple $(z_k, v_k, \varepsilon_k)$ satisfying (2.4) can be found by suitably applying the ACG method described in Subsection A.1 to subproblem (1.4).

We now formally describe S-IAIPAL.

**S-IAIPAL**

(0) Let scalars $\nu > 0$ and $\sigma \in (0, 1/\sqrt{2}]$, initial point $z_0 \in \mathcal{H}$, tolerance pair $(\hat{\rho}, \hat{\eta}) \in \Re_{++} \times \Re_{++}$, penalty parameter $c \geq 0$, and $\alpha > 0$ be given; set $k = 1$, $p_0 = 0$, and

$$C_1 = \frac{2\,(1+2\nu)^2}{1-\sigma^2}, \quad \lambda = \frac{1}{2m_f}, \quad \sigma_c = \min\left\{\frac{\nu}{\sqrt{\lambda L_c + 1}}, \sigma\right\}, \tag{2.6}$$
$$L_c = L_f + c\|A\|^2;$$

(1) use the ACG method described in Subsection A.1 with inputs

$$x_0 = z_{k-1}, \quad \tilde{\sigma} = \sigma_c, \quad (\widetilde{\mu}, \widetilde{M}) = (1/2, \lambda L_c + 1), \tag{2.7}$$

$$(\psi^{(s)}, \psi^{(n)}) = \left(\lambda[\mathcal{L}_c(\cdot, p_{k-1}) - h] + \frac{1}{2}\|\cdot - z_{k-1}\|^2,\ \lambda h\right) \tag{2.8}$$

to obtain a triple $(z_k, v_k, \varepsilon_k)$ satisfying (2.4) with $\tilde{\sigma} = \sigma_c$, and set

$$q_k = p_{k-1} + c(Az_k - b), \quad p_k = \begin{cases} q_k, & k \equiv 1 \bmod \lceil \alpha c \rceil, \\ p_{k-1}, & \text{otherwise;} \end{cases} \tag{2.9}$$

(2) compute $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ as

$$\hat{z}_k = \operatorname*{argmin}_u \left\{\langle \lambda\,[\nabla f(z_k) + A^* q_k] - r_k, u\rangle + \lambda h(u) + \frac{\lambda L_c + 1}{2}\|u - z_k\|^2\right\},$$

$$\hat{p}_k = p_{k-1} + c(A\hat{z}_k - b), \tag{2.10}$$

$$\hat{w}_k = \frac{1}{\lambda} r_k + \left(L_c + \frac{1}{\lambda} + cA^*A\right)(\hat{z}_k - z_k) + \nabla f(\hat{z}_k) - \nabla f(z_k), \tag{2.11}$$

where $r_k$ is as in (2.5); if $\|\hat{w}_k\| \leq \hat{\rho}$ and $\|A\hat{z}_k - b\| \leq \hat{\eta}$, then **stop with success** and output $(\hat{z}, \hat{p}, \hat{w}) = (\hat{z}_k, \hat{p}_k, \hat{w}_k)$;

(3) if $k \geq 2$ and

$$\Delta_k = \frac{\mathcal{L}_c(z_1, p_1) - \mathcal{L}_c(z_k, p_k)}{k-1} \leq \frac{\lambda \hat{\rho}^2}{2C_1}, \tag{2.12}$$

then **stop and declare** $c$ **small**;

(4) set $k \leftarrow k + 1$, and go to step 1.

We now make some trivial remarks about S-IAIPAL. First, it performs two types of iterations, namely, the outer ones indexed by $k$ and the ACG (or inner) ones performed during its calls to ACG in step 1. Second, the scalar $\lambda$ defined in step 0 ensures that the prox augmented Lagrangian subproblem (1.4) is strongly convex. Third, the scalars $\widetilde{M}$ and $\widetilde{\mu}$ in step 1 are the Lipschitz constant and the strong convexity parameter of $\nabla\psi_s$ and $\psi_n$, respectively. Fourth, the update formula (2.9) for the multiplier $p_k$ is the classical one where a full step is performed, i.e., no shrinking factor multiplying the term $c(Az_k - b)$ is included on it. Fifth, it follows immediately from (2.9) and (2.10) that

$$\hat{p}_k - p_k = cA(\hat{z}_k - z_k) \quad \forall k \equiv 1 \bmod \lceil \alpha c \rceil. \tag{2.13}$$

We next make some comments about the logical structure of S-IAIPAL. First, it is shown in Proposition 3.1 that every triple $(\hat{z}, \hat{p}, \hat{w}) = (\hat{z}_k, \hat{p}_k, \hat{w}_k)$ computed in step 2 satisfies the inclusion in (1.2), and hence is a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1.1) (see Definition 2.1) whenever S-IAIPAL stops successfully (see the condition for that to happen at the end of step 2). Second, in contrast to the $k$-th generated triple $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$, which is only used in step 2 to test for possible termination, the $k$-th generated quadruple $(z_k, p_k, v_k, \varepsilon_k)$ found in step 1 is not only used to compute the above triple but also to perform the next iteration. Third, Theorem 2.3(d) below shows that if the penalty parameter $c$ is sufficiently large at some iteration, then S-IAIPAL must successfully stop in its step 2. Finally, after the second iteration (and including it) of S-IAIPAL, inequality (2.12) is used to detect whether the penalty parameter $c$ is small, in which case S-IAIPAL stops in its step 3 with the declaration that $c$ is small. IAIPAL, which is discussed in the next subsection, then uses this information to increase $c$ and restart S-IAIPAL with the new value of $c$ and with the initial point $z_0$ either set to be the same as in the previous S-IAIPAL call, i.e., $z_0$ is kept constant (cold S-IAIPAL restart), or set to be equal to $z_k$, where $z_k$ is the iterate computed in step 1 of S-IAIPAL just before it declares $c$ small (warm S-IAIPAL restart).

The following result describes an upper bound on the number of iterations performed during each call to ACG in step 1 of S-IAIPAL.

PROPOSITION 2.2. *Each call to the ACG method in step 1 of S-IAIPAL performs at most*

$$\left\lceil 5 \left( \sqrt{\frac{2L_f}{m_f}} + \sqrt{\frac{c\|A\|^2}{m_f}} \right) \log_1^+ \mathcal{M}(c) \right\rceil \tag{2.14}$$

*ACG iterations, where $\mathcal{M}(c)$ is given by*

$$\mathcal{M}(c) = 2 \left[ \frac{3L_f}{m_f} + \frac{c\|A\|^2}{m_f} \right] \max\{\nu^{-1}, \sigma^{-1}\}. \tag{2.15}$$

*Proof.* First note that the respective definitions of $(\lambda, \sigma_c, L_c)$, $(\tilde{\sigma}, \widetilde{\mu}, \widetilde{M})$, and $\mathcal{A}_{\widetilde{\mu},\tilde{\sigma}}$ in (2.6), (2.7), and Proposition A.1, together with the bounds $\sigma_c < 1$ and $L_f/m_f \geq 1$ from the definition of $\sigma_c$ and **(B5)**, imply that

$$\mathcal{A}_{\widetilde{\mu},\sigma_c} = \frac{4(1+\sigma_c)^2}{\sigma_c^2} \leq \frac{16}{\sigma_c^2} \leq 8 \left( \frac{3L_f}{m_f} + \frac{c\|A\|^2}{m_f} \right) \max\{\nu^{-2}, \sigma^{-2}\},$$

$$\widetilde{M} - \widetilde{\mu} = \lambda L_c + \frac{1}{2} = \frac{L_f + c\|A\|^2}{2m_f} + \frac{1}{2} \leq \frac{1}{2} \left( \frac{2L_f}{m_f} + \frac{c\|A\|^2}{m_f} \right). \qquad \square$$

Hence, (2.14) follows from Proposition A.1, the above inequalities, the definition of $\mathcal{M}(c)$ in (2.15), and the fact that $\log_1^+(\cdot) \geq 1$.

The following quantities and constants will be used in the statement and proof of the main result of this subsection (Theorem 2.3 below).

$$C_2 := \frac{\sigma^2}{(1-\sigma)^2}, \quad C_3 := \frac{1+\nu}{1-\sigma}, \tag{2.16}$$

$$\phi_* := \inf_{z \in \Re^n} \phi(z), \quad \Delta\phi^* = \phi^* - \phi_*, \quad \bar{d} := \operatorname{dist}_{\partial\mathcal{H}}(\bar{z}), \tag{2.17}$$

$$\theta_A := \frac{\|A\|}{\sigma_A^+}, \quad \theta_D = \frac{D}{\bar{d}}, \tag{2.18}$$

$$\kappa_0 := 2(L_h + \nabla_f) + (C_2 + 4C_3)m_f D, \tag{2.19}$$

$$\kappa_1 := 2\sqrt{2C_1}\theta_A\theta_D\kappa_0, \qquad \kappa_2 := \left(\frac{5\|A\|\theta_A\theta_D\kappa_0}{2m_f}\right)^{1/2}, \tag{2.20}$$

where $\phi^*$ is as in (1.1), $C_1$ is as in (2.6), $D$ and $\nabla_f$ are as in **(B3)**, and $\bar{z}$ is as in **(B4)**. Note that $C_1$, $C_2$ and $C_3$ are constants depending only on the input parameters $\nu$ and/or $\sigma$ of S-IAIPAL. Moreover, the constants $\kappa_0$, $\kappa_1$ and $\kappa_2$ depend not only on the constants $C_1$, $C_2$ and $C_3$, but also on the constants $D$, $\|A\|$, $L_h$, $m_f$, $\nabla_f$, and the ones defined in (2.17) and (2.18), which are all associated with the instance of (1.1) under consideration. Constants $\kappa_1$ and $\kappa_2$ are in turn used to describe a threshold value $\bar{c}$ (see (2.24) below) such that if $c \geq \bar{c}$ then S-IAIPAL is guaranteed to terminate with a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1.1) (see statement (d) below).

Next we state the main result about S-IAIPAL, whose proof is given at the end of Section 3.

THEOREM 2.3. *Assume that $c \geq m_f/\|A\|^2$ and that conditions* **(B1)**–**(B5)** *hold. Then, the following statements about S-IAIPAL hold:*

*a) every iterate $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ with $k \geq 1$ satisfies $\hat{w}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^*\hat{p}_k$;*

*b) the number of outer iterations is bounded by*

$$T_0(\hat{\rho}) := \left\lceil 1 + \frac{12C_1 m_f\left(\Delta\phi^* + 2m_f D^2\right) + \kappa_1^2/4}{\hat{\rho}^2} \right\rceil, \tag{2.21}$$

*where $C_1$, $\Delta\phi^*$, $\theta_D$, $\kappa_1$, $D$, and $m_f$ are as in step 0 of S-IAIPAL, (2.17), (2.18), (2.20), **(B3)**, and **(B5)**, respectively; hence, the total number of ACG iterations is bounded by*

$$T_{ACG}(c, \hat{\rho}) := \left\lceil 5\left(\sqrt{\frac{2L_f}{m_f}} + \sqrt{\frac{c\|A\|^2}{m_f}}\right)\log_1^+ \mathcal{M}(c)\right\rceil T_0(\hat{\rho}), \tag{2.22}$$

*where $L_f$ and $\mathcal{M}(c)$ are as in **(B5)** and (2.15), respectively.*

*Moreover, if the penalty parameter $c$ satisfies*

$$c \geq \frac{\kappa_2^2 m_f}{\hat{\eta}\|A\|^2}, \quad c\lceil c\alpha\rceil \geq \frac{4m_f\kappa_1^2}{\hat{\rho}^2\|A\|^2} \tag{2.23}$$

*then the following statements also hold:*

*c) every iterate $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ with $k \geq 1$ satisfies $\|A\hat{z}_k - b\| \leq \hat{\eta}$;*

*d) S-IAIPAL stops successfully in step 2 with a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution $(\hat{z}, \hat{p}, \hat{w})$ of (1.1).*

We now make some remarks about Theorem 2.3 under the condition that the penalty parameter $c$ satisfies $\hat{c}_\alpha \leq c = \mathcal{O}(\hat{c}_\alpha)$ where

$$\hat{c}_\alpha = \hat{c}_\alpha(\hat{\rho}, \hat{\eta}) := \min\{c : c \text{ satisfies (2.23)}\}. \tag{2.24}$$

First, it follows from parts b) and d) that S-IAIPAL obtains a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1.1) in $\mathcal{O}\left(T_{ACG}(\hat{c}_\alpha, \hat{\rho})\right)$ ACG iterations, where $T_{ACG}(c, \hat{\rho})$ is as in (2.22). Second, under the reasonable assumption that right-hand-side of the second bound in (2.23) is $\Omega(1)$, it is easy to show that

$$\hat{c}_\alpha(\hat{\rho}, \hat{\eta}) = \Theta\left(\frac{\sqrt{m_f}}{\|A\|}\max\left\{\frac{\kappa_2^2\sqrt{m_f}}{\hat{\eta}\|A\|}, \frac{\kappa_1}{\hat{\rho}}\right\}\right) \quad \text{if } \alpha = \Theta(1), \tag{2.25}$$

$$\hat{c}_\alpha(\hat{\rho},\hat{\eta}) = \Theta\left(\frac{m_f}{\|A\|^2}\max\left\{\frac{\kappa_2^2}{\hat{\eta}},\frac{\kappa_1^2}{\hat{\rho}^2}\right\}\right) \quad \text{if } \alpha = \Theta(1/c). \tag{2.26}$$

This remark together with the fact that $T_0(\hat{\rho}) = \mathcal{O}(\hat{\rho}^{-2})$ then imply that the ACG iteration complexity of S-IAIPAL, up to a logarithmic term, is $\mathcal{O}(\hat{\rho}^{-5/2} + \hat{\rho}^{-2}\hat{\eta}^{-1/2})$ when $\alpha = \Theta(1)$, and $\mathcal{O}(\hat{\rho}^{-3} + \hat{\rho}^{-2}\hat{\eta}^{-1/2})$ when $\alpha = \Theta(1/c)$. Third, the number of iterations where S-IAIPAL performs a full multiplier update (i.e., $p_k = q_k$) is $\mathcal{O}(\hat{\rho}^{-2}[\alpha c]^{-1})$. In particular, if $\alpha = \Theta(1)$ and $\hat{\rho} = \hat{\eta}$, then the number of full multiplier updates is $\mathcal{O}(\hat{\rho}^{-1})$ when $c = \Theta(\hat{c}_\alpha)$ and is $\mathcal{O}(\hat{\rho}^{-2})$ when $c = \Theta(1)$. Fourth, since the threshold $\hat{c}_\alpha$ in (2.24) is not computable in practice, it is not clear how one can choose a penalty parameter $c$ such that $\hat{c}_\alpha \le c = \mathcal{O}(\hat{c}_\alpha)$.

The next subsection presents IAIPAL which repeatedly invokes S-IAIPAL with increasing penalty parameter values until a $(\hat{\rho},\hat{\eta})$–approximate solution of (1.1) is obtained. Moreover, it is shown that, up to a logarithmic term, the overall number of ACG iterations performed by this scheme is the same as the one of S-IAIPAL under the condition $\hat{c}_\alpha \le c = \mathcal{O}(\hat{c}_\alpha)$.

**2.3. The IAIPAL method.** This subsection describes the IAIPAL method and establishes its ACG iteration-complexity.

The statement of IAIPAL below makes use of S-IAIPAL presented in Subsection 2.2. More specifically, it consists of repeatedly invoking S-IAIPAL with $c = c_\ell := c_1\tau^{\ell-1}$ where $c_1$ is an initial choice for the penalty parameter, $\tau > 1$, and $\ell$ is the S-IAIPAL call count.

**IAIPAL**

(0) Let a quadruple of scalars $(\nu,\sigma,\tau) \in \Re_{++}\times(0,1/\sqrt{2}]\times(1,+\infty)$ and a pair of tolerances $(\hat{\rho},\hat{\eta}) \in \Re_{++}\times\Re_{++}$ be given; choose $c_1 > 0$ and set $\ell \leftarrow 1$;
(1) choose an initial point $z_0^{(\ell)} \in \mathcal{H}$ and some $\alpha_\ell \in \Re_{++}$; call S-IAIPAL with inputs $z_0 = z_0^{(\ell)}$, $\nu$, $\sigma$, $\hat{\rho}$, $\hat{\eta}$, $c = c_\ell$, and $\alpha = \alpha_\ell$;
(2) if S-IAIPAL successfully stops with a triple $(\hat{z},\hat{p},\hat{w})$, then output this triple and stop; otherwise, set $c_{\ell+1} \leftarrow \tau c_\ell$, set $\ell \leftarrow \ell+1$, and return to step 1.

We now make some remarks about IAIPAL. First, the initial point $z_0^{(\ell)}$ chosen in step 1 can either be the same point (cold start) across all S-IAIPAL calls or a varying point. In the latter case, a simple approach (warm start) is to choose $z_0^{(\ell)}$ as the last iterate computed in the most recent call to S-IAIPAL. Second, every outer iteration within the $\ell$-th S-IAIPAL call uses the penalty parameter $c_\ell = c_1\tau^{\ell-1}$. Third, if $\ell$-th S-IAIPAL call does not successfully stop in step 2 or, equivalently, declares $c_\ell$ small in step 3 of S-IAIPAL, then the next penalty parameter $c_{\ell+1}$ is increased by a multiplicative factor $\tau > 1$. Finally, $\alpha_\ell$ can be chosen as a constant in every execution of step 1, or it can change. For example, choosing $\alpha_\ell = 1/c_\ell$ guarantees that a Lagrange multiplier update is performed at every outer iteration of an S-IAIPAL call.

The following result establishes the overall ACG iteration-complexity for IAIPAL to obtain a $(\hat{\rho},\hat{\eta})$-approximate stationary solution of (1.1).

THEOREM 2.4. *Assume that conditions* **(B1)**–**(B5)** *of Subsection 2.1 hold and define the scalar*

$$\hat{c}(\hat{\rho},\hat{\eta}) := \sup_{\ell\ge 1}\hat{c}_{\alpha_\ell}(\hat{\rho},\hat{\eta}), \tag{2.27}$$

*where $\hat{c}_{\alpha_\ell}(\cdot,\cdot)$ is as in (2.24). Then, IAIPAL obtains a $(\hat{\rho},\hat{\eta})$-approximate stationary*

*solution* $(\hat{z}, \hat{p}, \hat{w})$ *of problem* (1.1) *in*

$$\mathcal{O}\left(T_{ACG}(\hat{c}(\hat{\rho},\hat{\eta})+c_1,\hat{\rho})\cdot\log_1^+\left\{\frac{\hat{c}(\hat{\rho},\hat{\eta})}{c_1}\right\}\right) \tag{2.28}$$

*ACG iterations, where* $c_1$ *is the initial penalty parameter in IAIPAL,* $\kappa_1$ *and* $\kappa_2$ *are as in* (2.20)*, and* $T_{ACG}(\cdot,\cdot)$ *is as in* (2.22)*.*

*Proof.* First note that the $\ell$-th loop of IAIPAL invokes S-IAIPAL with penalty parameter $c_\ell = \tau^{\ell-1}c_1$, for every $\ell \geq 1$. It is easy to see that if IAIPAL stops in its first call to S-IAIPAL, then the statement of the theorem follows trivially in view of the stopping criterion in step 2 of IAIPAL and Theorem 2.3(b). Suppose then that IAIPAL calls S-IAIPAL more than once and let $\hat{c} = \hat{c}(\hat{\rho},\hat{\eta})$. Defining the integer

$$\bar{\ell} := \min\{\ell : c_\ell \geq \hat{c}\}, \tag{2.29}$$

it follows from Theorem 2.3(d) that a $(\hat{\rho},\hat{\eta})$-approximate solution of (1.1) is obtained in at most $\bar{\ell} \geq 2$ calls to S-IAIPAL. In view of the minimality in (2.29) and the penalty update rule in step 2 of IAIPAL, we have $c_{\bar{\ell}} \leq \tau\hat{c}$ and, hence,

$$\bar{\ell} = \log_\tau(\tau^{\bar{\ell}}) = \log_\tau\frac{\tau c_{\bar{\ell}}}{c_1} \leq \log_\tau\frac{\tau^2\hat{c}}{c_1} = 2+\log_\tau\frac{\hat{c}}{c_1}. \tag{2.30}$$

Combining (2.30), the fact that $T_{ACG}(\hat{c},\hat{\rho}) \geq T_{ACG}(\hat{c}_{\alpha_\ell},\hat{\rho})$ for $\ell \geq 1$, and Theorem 2.3(b), we conclude that the number of ACG iterations of IAIPAL is on the same order of magnitude as in (2.28). ∎

We now make some remarks about Theorem 2.4. First, it is easy to see that for fixed $(\hat{\rho},\hat{\eta})$, it holds that $\sup_{\alpha>0}\hat{c}_\alpha(\hat{\rho},\hat{\eta})$ is finite and, hence, $\hat{c}$ in (2.27) is also finite. Second, its iteration-complexity does not depend on how $z_0$ is selected in step 0. As a consequence, it applies to both the cold start and the warm start approaches mentioned above. Third, it follows from Theorem 2.4 that the total number of ACG iterations of IAIPAL is, up to a logarithmic term, the same as that of S-IAIPAL with penalty parameter $c$ such that $\hat{c}(\hat{\rho},\hat{\eta}) \leq c = \mathcal{O}(\hat{c}(\hat{\rho},\hat{\eta}))$.

The next result describes (2.28) only in terms of $(\hat{\rho},\hat{\eta})$ for two choices of $\alpha_\ell$.

COROLLARY 2.5. *Assume that conditions* **(B1)**–**(B5)** *of Subsection 2.1 hold and that* $\max\{c_1, c_1^{-1}\} = \mathcal{O}(\hat{c}(\hat{\rho},\hat{\eta}))$*, where* $\hat{c}(\cdot,\cdot)$ *is as in* (2.27)*. Then, IAIPAL obtains a* $(\hat{\rho},\hat{\eta})$*-approximate stationary solution of problem* (1.1) *in a number of ACG iterations/resolvent evaluations bounded, up to a logarithmic term, by*

$$\mathcal{O}(\hat{\rho}^{-5/2}+\hat{\rho}^{-2}\hat{\eta}^{-1/2}) \quad \text{if } \alpha_\ell = \Theta(1), \tag{2.31}$$

$$\mathcal{O}(\hat{\rho}^{-3}+\hat{\rho}^{-2}\hat{\eta}^{-1/2}) \quad \text{if } \alpha_\ell = \Theta(1/c_\ell). \tag{2.32}$$

*Consequently, if IAIPAL performs a multiplier update at every outer iteration of S-IAIPAL, i.e., choose* $\alpha_\ell = 1/c_\ell$ *in step 1 of IAIPAL, then its ACG iteration complexity is as in* (2.32)*.*

*Proof.* This follows immediately from Theorem 2.4, the definition of $T_{ACG}$ in (2.22), and (2.25)–(2.26). ∎

## 3. Proof of Theorem 2.3.

The goal of this section is to provide the proof of Theorem 2.3 which describes the main properties of S-IAIPAL.

We start by motivating the results developed in this section. A major part of our effort lies in showing that the residual and the feasibility gap sequences $\{\|\hat{w}_k\|\}$ and $\{\|A\hat{z}_k - b\|\}$ generated by S-IAIPAL with penalty parameter $c$ satisfy

$$\min_{i\le k} \|\hat{w}_i\| = \mathcal{O}\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{\alpha}c}\right), \quad \|A\hat{z}_k - b\| = \mathcal{O}\left(\frac{1}{c}\right), \quad \forall k \ge 2. \tag{3.1}$$

Observe that (3.1) implies that there exists a range of sufficiently large values of $c$ satisfying $c^{-1} = \mathcal{O}(\min\{\hat{\rho}\sqrt{\alpha}, \hat{\eta}\})$ and such that S-IAIPAL finds a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1.1) in $\mathcal{O}(\hat{\rho}^{-2})$ S-IAIPAL iterations. Using this observation together with Proposition 2.2, it is now easy to see that there exists a significantly large range of $c$'s for which the total number of ACG iterations performed by S-IAPIAL is $\mathcal{O}(\hat{\rho}^{-5/2} + \hat{\rho}^{-2}\hat{\eta}^{-1/2})$, up to a multiplicative logarithmic term. Lemma 3.3(b) below establishes a key inequality towards proving the first relation in (3.1), and the paragraph following this lemma outlines how this inequality is used to establish (3.1).

The first technical result below describes some important properties about the sequence $\{(\hat{z}_k, \hat{p}_k, \hat{w}_k)\}$ computed in step 2 of S-IAIPAL as well as other related sequences which are also used in the analysis of S-IAIPAL.

PROPOSITION 3.1. *The following statements hold:*

*a) the triple $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ generated in step 2 of S-IAIPAL and the residual $r_k$ defined in (2.5) satisfy*

$$\hat{w}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^*\hat{p}_k, \tag{3.2}$$

$$\lambda\|\hat{w}_k\| \le (1+2\nu)\,\|r_k\|, \qquad \|\hat{z}_k - z_k\| \le \frac{\nu}{2(\lambda L_c + 1)}\|r_k\|, \tag{3.3}$$

*where $\nu$ and $L_c$ are as in (2.6);*

*b) the quadruple $(z_k, p_k, w_k, \varepsilon_k)$ where $w_k$ is defined as*

$$w_k := \frac{1}{\lambda}\left[(\lambda L_c + 1)(z_k - \hat{z}_k) + r_k\right] \tag{3.4}$$

*and $(z_k, q_k, \varepsilon_k)$ is computed in step 1 of S-IAIPAL, satisfies*

$$w_k \in \nabla f(z_k) + \partial_{(\lambda^{-1}\varepsilon_k)} h(z_k) + A^* q_k, \tag{3.5}$$

$$\lambda\|w_k\| \le (1+\nu)\,\|r_k\|, \qquad \varepsilon_k \le \frac{\sigma_c^2\|r_k\|^2}{2} \tag{3.6}$$

*where $\sigma_c$ is as in (2.6).*

*Proof.* First note that the last inequality in (3.6) follows immediately from the inequality in (2.4) with $\tilde{\sigma} = \sigma_c$ in view of step 1. Note also that the quantities $(\tilde{g}, \tilde{h})$, $(z, \varepsilon)$, and $\tilde{L}$ defined as

$$\tilde{g} := \lambda[\mathcal{L}_c(\cdot, p_{k-1}) - h] - \langle v_k, \cdot - z_k\rangle + \frac{1}{2}\|\cdot - z_{k-1}\|^2, \quad \tilde{h} := \lambda h, \tag{3.7}$$

$$(z, \varepsilon) := (z_k, \varepsilon_k), \quad \tilde{L} := \lambda L_c + 1 \tag{3.8}$$

satisfy the assumptions of Lemma A.3, in view of **(B2)**, **(B5)**, (1.3), (1.8), (1.9), (2.1), and the inclusion in (2.4). Observe also that (2.5), (2.9), and (3.7) imply that $\tilde{z}$ and $\widetilde{w}$ in (A.5) are equal to $\hat{z}_k$ and $(\lambda L_c + 1)(z_k - \hat{z}_k)$, respectively, and

$\nabla \tilde{g}(z_k) = \lambda[\nabla f(z_k) + A^* q_k] - r_k$. Hence, it follows from the conclusion of Lemma A.3 that

$$(\lambda L_c + 1)(z_k - \hat{z}_k) + r_k \in \lambda[\nabla f(z_k) + A^* q_k] + \partial(\lambda h)(\hat{z}_k), \tag{3.9}$$

$$(\lambda L_c + 1)(z_k - \hat{z}_k) + r_k \in \lambda[\nabla f(z_k) + A^* q_k] + \partial_{\varepsilon_k}(\lambda h)(z_k), \tag{3.10}$$

$$(\lambda L_c + 1)\|(z_k - \hat{z}_k)\| \leq \sqrt{2(\lambda L_c + 1)\varepsilon_k}. \tag{3.11}$$

Hence, inclusion (3.2) follows from (2.11), (2.13), (3.9), and a well-known property of the $\varepsilon$-subdifferential of a function which follows directly from its definition (1.8). Moreover, inclusion (3.5) follows immediately from (3.4) and (3.10). The first inequality in (3.6) follows from (3.4), the Cauchy-Schwarz inequality, (3.11), the last inequality in (3.6), and the definition of $\sigma_c$ in (2.6). Now, (2.1), (2.11), (3.4), (3.11), the definition of $L_c$ in (2.6), and the Cauchy-Schwarz inequality, imply that

$$\lambda\|\hat{w}_k\| \leq \|\lambda w_k\| + \lambda(L_f + c\|A\|^2)\|\hat{z}_k - z_k\| \leq \|\lambda w_k\| + \lambda\sqrt{2(\lambda L_c + 1)\varepsilon_k}.$$

The first inequality in (3.3) then follows from the above inequalities together with (3.6) and the definition of $\sigma_c$ in (2.6). Finally, the second inequality in (3.3) follows immediately from (3.11), the last inequality in (3.6), and the definition of $\sigma_c$ in (2.6).□

We now make two remarks about Proposition 3.1. First, the residual $w_k$ in (3.4) does not appear in the description of S-IAIPAL (and hence IAIPAL), but it plays an important role in its analysis. More specifically, the residual pair $(w_k, \varepsilon_k)$, and the corresponding bounds developed for it in (3.6), play a crucial role in proving that the sequence $\{p_k\}$ of Lagrange multipliers is bounded. Second, the right hand sides of the inequalities in (3.3) and (3.6) are all expressed in terms of $\|r_k\|$ since a substantial part of our analysis will concentrate on deriving suitable bounds for it, and hence for the quantities which are bounded in (3.3) and (3.6).

The following technical result derives an estimate on $\{\|r_k\|\}$ in terms of the variation of the augmented Lagrangian function along the sequence $\{(z_k, p_k)\}$ and the variation of the sequence of Lagrangian multipliers $\{p_k\}$.

LEMMA 3.2. *Let $\{(z_k, p_k, v_k, \varepsilon_k)\}$ be generated by S-IAIPAL, let $\{r_k\}$ be as in (2.5), and define $\{\Delta p_k\}$ as*

$$\Delta p_k := p_k - p_{k-1}, \qquad \forall k \geq 1. \tag{3.12}$$

*Then, the following inequality holds for every $k \geq 1$:*

$$\|r_k\|^2 \leq \frac{2\lambda}{1 - \sigma_c^2}\left(\mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_k) + \frac{1}{c}\|\Delta p_k\|^2\right). \tag{3.13}$$

*Proof.* In view of the update rule for $p_k$ given in step 1 of S-IAIPAL and the definitions of $\mathcal{L}_c$ and $\Delta p_k$ given in (1.3) and (3.12), respectively, we have

$$\mathcal{L}_c(z_k, p_k) - \mathcal{L}_c(z_k, p_{k-1}) = \langle \Delta p_k, Az_k - b\rangle = \frac{1}{c}\|\Delta p_k\|^2, \tag{3.14}$$

where the last identity follows from the fact that $\Delta p_k = 0$ when $k \not\equiv 1 \bmod \lceil \alpha c \rceil$ and $Az_k - b = c^{-1}\Delta p_k$ when $k \equiv 1 \bmod \lceil \alpha c \rceil$. Now, it follows from (1.8), (2.4), and (2.5) that

$$\lambda\mathcal{L}_c(z_k, p_{k-1}) - \lambda\mathcal{L}_c(z_{k-1}, p_{k-1}) \leq -\frac{1}{2}\|z_k - z_{k-1}\|^2 + \langle v_k, z_k - z_{k-1}\rangle + \varepsilon_k$$

$$= -\frac{1}{2}\|v_k + z_{k-1} - z_k\|^2 + \frac{\|v_k\|^2}{2} + \varepsilon_k \leq -\frac{1-\sigma_c^2}{2}\|r_k\|^2,$$

which implies that

$$\frac{1-\sigma_c^2}{2\lambda}\|r_k\|^2 \leq \mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_{k-1}).$$

The inequality in (3.13) then follows by combining the latter inequality with (3.14). ∎

Recall that Proposition 3.1(a) implies that the triple $(\hat{z}, \hat{p}, \hat{w}) = (\hat{z}_k, \hat{p}_k, \hat{w}_k)$ satisfies the inclusion in (1.2). The following technical result gives a preliminary bound on $\|\hat{w}_k\|$ and establishes the key inequality mentioned in the second paragraph of this section.

LEMMA 3.3. *Consider the sequences $\{(z_k, p_k, v_k, \varepsilon_k)\}$ and $\{(\hat{z}_k, \hat{p}_k, \hat{w}_k)\}$ generated by S-IAIPAL and let $C_1$, $\Delta_k$, and $\Delta p_k$ be as in (2.6), (2.12), and (3.12), respectively. Then, the following statements hold:*

*a) for every $k \geq 1$, we have*

$$\|\hat{w}_k\|^2 \leq \frac{C_1}{\lambda}\left[\mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_k) + \frac{1}{c}\|\Delta p_k\|^2\right]; \tag{3.15}$$

*b) for every $k \geq 2$, we have*

$$\min_{2\leq i\leq k}\|\hat{w}_i\|^2 \leq \frac{C_1}{\lambda}\left[\Delta_k + \frac{1}{k-1}\sum_{i=2}^{k}\frac{\|\Delta p_i\|^2}{c}\right]. \tag{3.16}$$

*Proof.* a) It follows from Proposition 3.1(a) that the triple $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ computed in step 2 of S-IAIPAL satisfies, in particular, the first inequality in (3.3). This conclusion together with inequality (3.13) then imply that

$$\|\hat{w}_k\|^2 \leq \frac{(1+2\nu)^2\|r_k\|^2}{\lambda^2} \leq \frac{2(1+2\nu)^2}{\lambda(1-\sigma_c^2)}\left[\mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_k) + \frac{1}{c}\|\Delta p_k\|^2\right],$$

and hence that (3.15) holds, in view of the definition of $C_1$ and the fact $\sigma_c \leq \sigma$, see (2.6).

b) Summing inequality (3.15) from $k = 2$ to $k = k$, and using the definition of $\Delta_k$ given in (2.12), we obtain

$$(k-1)\min_{2\leq i\leq k}\|\hat{w}_i\|^2 \leq \sum_{i=2}^{k}\|\hat{w}_i\|^2 \leq \frac{C_1}{\lambda}\left[(k-1)\Delta_k + \sum_{i=2}^{k}\frac{\|\Delta p_i\|^2}{c}\right]. \qquad ∎$$

We now outline how (3.16), together with some technical results below, can be used to establish the first bound in (3.1). Bound (3.16) on $\min_{i\leq k}\|\hat{w}_i\|^2$ is the sum of several terms, one of which depends on $\Delta_k$. Now, Lemmas 3.5 and 3.6 show that $\Delta_k$ is $\mathcal{O}([1+\|p_k\|^2]k^{-1})$. Moreover, with the help of Lemmas 3.7–3.11, Proposition 3.12 establishes that $\|p_k\|$ is bounded by a constant independent of $c$. Note that (3.16), the above two observations, and the update rule (1.5) then imply that $\min_{i\leq k}\|\hat{w}_i\|^2$ behaves as $\mathcal{O}(k^{-1} + \alpha^{-1}c^{-2})$ and, hence, that the first relation in (3.1) holds.

The next result, whose proof can be found in [30, Lemma A.3], will be used in the proof of Lemma 3.5.

LEMMA 3.4. *Let proper function* $\tilde{\phi} : \Re^n \to (-\infty, \infty]$, *scalar* $\tilde{\sigma} \in (0,1)$ *and* $(z_0, z_1) \in \Re^n \times \operatorname{dom} \tilde{\phi}$ *be given, and assume that there exists* $(v_1, \varepsilon_1)$ *such that*

$$v_1 \in \partial_{\varepsilon_1} \left( \tilde{\phi} + \frac{1}{2} \| \cdot - z_0 \|^2 \right) (z_1), \quad \|v_1\|^2 + 2\varepsilon_1 \leq \tilde{\sigma}^2 \|v + z_0 - z_1\|^2. \tag{3.17}$$

*Then, for every* $z \in \Re^n$ *and* $s > 0$, *we have*

$$\tilde{\phi}(z_1) + \frac{1}{2} \left[ 1 - \tilde{\sigma}^2 (1 + s^{-1}) \right] \|v_1 + z_0 - z_1\|^2 \leq \tilde{\phi}(z) + \frac{s+1}{2} \|z - z_0\|^2.$$

The following technical result shows that $\mathcal{L}_c(z_1, p_1)$ can be majorized by a scalar which does not depend on $c$. This fact, which is not immediately apparent from the definition of $\mathcal{L}_c(\cdot, \cdot)$, plays an important role in showing that S-IAIPAL or IAIPAL can start from an arbitrary (and hence infeasible) point in $\mathcal{H}$.

LEMMA 3.5. *The first quadruple* $(z_1, p_1, v_1, \varepsilon_1)$ *generated by S-IAIPAL satisfies*

$$\mathcal{L}_c(z_1, p_1) \leq 3 \left( \Delta\phi^* + 2 m_f D^2 \right) + \phi_*, \tag{3.18}$$

*where* $\phi_*$ *and* $\Delta\phi^*$ *are as in* (2.17).

*Proof.* The fact that $(z_1, v_1, \varepsilon_1)$ satisfies (2.4) with $k = 1$ and $\tilde{\sigma} = \sigma_c$, Lemma 3.4 with $s = 1$ and $\tilde{\phi} = \lambda \mathcal{L}_c(\cdot, p_0)$, and condition **(B3)**, imply that for every $z \in \mathcal{H}$,

$$\lambda \mathcal{L}_c(z_1, p_0) + \frac{1 - 2\sigma_c^2}{2} \|r_1\|^2 \leq \lambda \mathcal{L}_c(z, p_0) + \|z - z_0\|^2 \leq \lambda \mathcal{L}_c(z, p_0) + D^2,$$

where $r_1$ is as in (2.5). Using the definitions of $\phi^*$ and $\lambda$ given in (1.1) and (2.6), respectively, the fact that $1 - 2\sigma_c^2 \geq 1 - 2\sigma^2 \geq 0$ due to the definitions of $\sigma$ and $\sigma_c$ in step 0 of S-IAIPAL, and the fact that the definition of $\mathcal{L}_c$ in (1.3) implies that $\mathcal{L}_c(z, p_0) = (f + h)(z)$ for every $z \in \mathcal{F} := \{z \in \mathcal{H} : Az = b\}$, we then conclude from the above inequality, as $z$ varies in $\mathcal{F}$, that

$$\mathcal{L}_c(z_1, p_0) \leq \phi^* + 2 m_f D^2.$$

The above inequality together with the fact that $p_0 = 0$, (2.9) with $k = 1$, and the definitions of $\mathcal{L}_c$ and $\phi_*$ given in (1.3) and (2.17), respectively, then imply that

$$\begin{aligned} \mathcal{L}_c(z_1, p_1) &= \mathcal{L}_c(z_1, p_0) + c\|Az_1 - b\|^2 = 3\mathcal{L}_c(z_1, p_0) - 2(f + h)(z_1) \\ &\leq 3(\phi^* + 2 m_f D^2) - 2\phi_*, \end{aligned}$$

which proves (3.18) in view of the definition of $\Delta\phi^*$. ∎

The following technical result shows that $\Delta_k = \mathcal{O}([1 + c^{-1}\|p_k\|^2] k^{-1})$.

LEMMA 3.6. *Let* $\{(z_k, p_k)\}$ *be generated by S-IAIPAL and consider* $\{\Delta_k\}$ *as in* (2.12). *Then, the following statements hold:*

*a) for every* $k \geq 1$, *we have*

$$\mathcal{L}_c(z_k, p_k) + \frac{\|p_k\|^2}{2c} \geq \phi_*, \tag{3.19}$$

*where* $\phi_*$ *is as in* (2.17);

*b) for every $k \geq 2$, we have*

$$\Delta_k \leq \frac{1}{k-1}\left[3\left(\Delta\phi^* + 2m_f D^2\right) + \frac{\|p_k\|^2}{2c}\right], \tag{3.20}$$

*where $\Delta\phi^*$ is as in (2.17).*

*Proof.* (a) Using the definitions of $\mathcal{L}_c$ and $\phi_*$ given in (1.3) and (2.17), respectively, we have

$$\begin{aligned}\mathcal{L}_c(z_k, p_k) &= (f+h)(z_k) + \langle p_k, Az_k - b\rangle + \frac{c}{2}\|Az_k - b\|^2 \\ &\geq \phi_* + \frac{1}{2}\left\|\frac{p_k}{\sqrt{c}} + \sqrt{c}(Az_k - b)\right\|^2 - \frac{1}{2c}\|p_k\|^2,\end{aligned}$$

and hence that (3.19) holds.

(b) This statement follows from (3.18), (3.19), and the definition of $\Delta_k$ in (2.12). ∎

The next technical results (i.e., Lemmas 3.7–3.11) develop the necessary tools for showing in Proposition 3.12 that the sequence $\{p_k\}$ is bounded. The first one gives some straightforward bounds among the different quantities involved in the analysis of S-IAIPAL.

LEMMA 3.7. *Let $\{(z_k, p_k, v_k, \varepsilon_k)\}$ be generated by S-IAIPAL and let $\{r_k\}$ be as in (2.5). Then, the following inequalities hold for every $k \geq 1$,*

$$\|r_k\| \leq \frac{D}{1-\sigma}, \quad \|v_k\|^2 \leq C_2 D^2, \quad \varepsilon_k \leq \frac{C_2 D^2}{2}, \tag{3.21}$$

*where $D$ is as in* **(B3)** *and $\sigma$ is as in step 0 of S-IAIPAL, and $C_2$ is as in (2.16).*

*Proof.* First note that, in view of step 1 of S-IAIPAL, the tuples $(\lambda, z_{k-1}, p_{k-1})$ and $(z_k, v_k, \varepsilon_k)$ satisfy (2.4). Hence, using the inequality in (2.4), the definition of $r_k$ given in (2.5), the triangle inequality, the first condition in **(B3)**, and the fact that $\sigma_c \leq \sigma$, we have

$$\|r_k\| - D \leq \|r_k\| - \|z_k - z_{k-1}\| \leq \|v_k\| \leq \sigma\|r_k\|, \qquad \varepsilon_k \leq \frac{\sigma^2\|r_k\|^2}{2}. \tag{3.22}$$

The first inequality in (3.21) immediately follows from the first setting of inequalities in (3.22). The last two inequalities in (3.21) follow from the first inequality in (3.21), the last two inequalities in (3.22) and the definition of $C_2$ in (2.16). ∎

The following basic result is used in Lemma 3.9. Its proof can be found, for instance, in [8, Lemma A.4]. Recall that $\sigma_A^+$ denotes the smallest positive singular value of a nonzero linear operator $A$ .

LEMMA 3.8. *Let $A : \Re^n \to \Re^l$ be a nonzero linear operator. Then, $\sigma_A^+\|u\| \leq \|A^*u\|$, for every $u \in A(\Re^n)$.*

The next result defines a slack $\xi_k \in \partial_{(\lambda^{-1}\varepsilon_k)}h(z_k)$ which realizes the inclusion in (3.5) and gives a preliminary bound on $\|p_k\|$ in terms of $\|\xi_k\|$.

LEMMA 3.9. *Consider the sequence $\{(z_k, q_k, v_k, \varepsilon_k)\}$ generated by S-IAIPAL and the sequence $\{w_k\}$ as in (3.4), and define*

$$\xi_k := w_k - \nabla f(z_k) - A^*q_k \tag{3.23}$$

*for every $k \geq 1$. Then, the following statements hold:*

*a) for every $k \geq 1$, we have*

$$\xi_k \in \partial_{(\lambda^{-1}\varepsilon_k)} h(z_k), \qquad \|w_k\| \leq \frac{C_3 D}{\lambda} \tag{3.24}$$

*where $D$ is as in* **(B3)** *and $C_3$ is as in (2.16);*

*b) for every $k \geq 1$, we have*

$$\sigma_A^+ \|q_k\| \leq \|\xi_k\| + \nabla_f + \frac{C_3 D}{\lambda}, \tag{3.25}$$

*where $\nabla_f$ is as in* **(B3)**.

*Proof.* (a) The inclusion in (3.24) follows from (3.5) and the definition of $\xi_k$ in (3.23). The inequality in (3.24) follows from the first inequalities in (3.6) and (3.21), and the definitions of $\sigma_c$ and $C_3$ in (2.6) and (2.16), respectively.

(b) Using **(B4)**, the fact that $p_0 = 0$ together with the update formula for $q_k$ and $p_k$, it is easy to see that $\{q_k\} \subset A(\Re^n)$. Using Lemma 3.8, relation (3.23), the triangle inequality, **(B3)**, and the inequality in (3.24), we conclude that

$$\sigma_A^+ \|q_k\| \leq \|A^* q_k\| \leq \|\xi_k\| + \|\nabla f(z_k)\| + \|w_k\| \leq \|\xi_k\| + \nabla_f + \frac{C_3 D}{\lambda}, \tag{3.26}$$

and, hence, that (3.25) holds. ∎

The next technical result essentially allows us to obtain a preliminary bound on $\|\xi_k\|$ under assumption **(B4)**. It is worth mentioning its proof is based on a key inequality that appears in the proof of Lemma 3 of [26].

LEMMA 3.10. *Let $h$ be a function as in* **(B2)**. *Then, for every $z, z' \in \mathcal{H}$, $\varepsilon > 0$, and $\xi \in \partial_\varepsilon h(z)$, we have*

$$\|\xi\| \mathrm{dist}_{\partial\mathcal{H}}(z') \leq (\mathrm{dist}_{\partial\mathcal{H}}(z') + \|z - z'\|) L_h + \langle \xi, z - z' \rangle + \varepsilon,$$

*where $\partial\mathcal{H}$ denotes the boundary of $\mathcal{H}$.*

*Proof.* Let $\varepsilon > 0$, $z, z' \in \mathcal{H}$ and $\xi \in \partial_\varepsilon h(z)$ be given. It follows from the Lipchitz continuity of $h$ in (**B2**) combined with the equivalence between (a) and (d) of Lemma A.2 that there exist $\xi_1 \in \bar{B}(0, L_h)$ and $\xi_2 \in N_{\mathcal{H}}^{\varepsilon}(z)$ such that $\xi = \xi_1 + \xi_2$. Clearly, it follows from the definitions of $\bar{B}(0, L_h)$ and $N_{\mathcal{H}}^{\varepsilon}(z)$ in Subsection 1.1 that

$$\|\xi_1\| \leq L_h, \quad \mathcal{H} \subset H_- := \{u \in \Re^n : \langle \xi_2, u - z \rangle - \varepsilon \leq 0\}.$$

Using the last inclusion and the fact that $z' \in \mathcal{H}$, we easily see that

$$\mathrm{dist}_{\partial\mathcal{H}}(z') \|\xi_2\| \leq \mathrm{dist}_{\partial H_-}(z') \|\xi_2\| = \langle \xi_2, z - z' \rangle + \varepsilon.$$

The last inequality, the fact that $\xi = \xi_1 + \xi_2$, the triangle inequality, and the Cauchy-Schwarz inequality, then imply that

$$\begin{aligned} \mathrm{dist}_{\partial\mathcal{H}}(z') \|\xi\| &\leq \mathrm{dist}_{\partial\mathcal{H}}(z') \|\xi_1\| + \mathrm{dist}_{\partial\mathcal{H}}(z') \|\xi_2\| \leq \mathrm{dist}_{\partial\mathcal{H}}(z') \|\xi_1\| + \langle \xi_2, z - z' \rangle + \varepsilon \\ &= \mathrm{dist}_{\partial\mathcal{H}}(z') \|\xi_1\| - \langle \xi_1, z - z' \rangle + \langle \xi, z - z' \rangle + \varepsilon \\ &\leq (\mathrm{dist}_{\partial\mathcal{H}}(z') + \|z - z'\|) \|\xi_1\| + \langle \xi, z - z' \rangle + \varepsilon, \end{aligned}$$

which combined with the fact that $\|\xi_1\| \leq L_h$ shows that the conclusion of the lemma holds. ∎

The next lemma presents some important technical inequalities using the bounds in Lemma 3.10 and Lemma 3.9(b).

LEMMA 3.11. *The iterates $\{(p_k, q_k, z_k)\}$ generated by S-IAIPAL satisfy:*
*a)* $\bar{d}\sigma_A^+\|q_k\| \leq D\kappa_0 - \langle q_k, Az_k - b\rangle$ *for every* $k \geq 1$*;*
*b)* $c\|Az_k - b\| \leq \theta_D\kappa_0(\sigma_A^+)^{-1} + \|p_{k-1}\|$ *for every* $k \geq 1$*;*
*c)* $c^{-1}\|p_k\|^2 + \bar{d}\sigma_A^+\|p_k\| \leq c^{-1}\langle p_k, p_{k-1}\rangle + D\kappa_0$ *for every* $k \equiv 1 \bmod k_c$*;*
*where* $\sigma_A^+$ *is defined in Subsection 1.1 and* $\bar{d}$*,* $\theta_D$*, and* $\kappa_0$ *are as in* (2.17)*,* (2.18)*, and* (2.19)*, respectively.*

*Proof.* (a) Let $\{\xi_k\}$ be as in (3.23). Using (3.21), (3.24), **(B3)**, the Cauchy-Schwarz and triangle inequalities, and the fact that $\lambda = 1/(2m_f)$ and $\|z_k - \bar{z}\| \leq D$, we first have that

$$\begin{aligned}\langle \xi_k, z_k - \bar{z}\rangle + 2m_f\varepsilon_k &\overset{(3.23)}{=} \langle w_k - \nabla f(z_k) - A^*q_k, z_k - \bar{z}\rangle + 2m_f\varepsilon_k \\ &\leq -\langle A^*q_k, z_k - \bar{z}\rangle + \|z_k - \bar{z}\|\left(\|w_k\| + \|\nabla f(z_k)\|\right) + 2m_f\varepsilon_k \\ &\leq -\langle q_k, Az_k - b\rangle + D\left(\nabla_f + [2C_3 + C_2]\, m_f D\right).\end{aligned}$$

Now, recall that $\bar{d} = \operatorname{dist}_{\partial\mathcal{H}}(\bar{z})$ and note that $\xi_k \in \partial_{(\lambda^{-1}\varepsilon_k)}h(z_k)$ for every $k \geq 1$, in view of (2.17) and Lemma 3.9(a), respectively. Hence, using the above technical bound, Lemma 3.9(b), Lemma 3.10 with $(\xi, z, z', \varepsilon) = (\xi_k, z_k, \bar{z}, \lambda^{-1}\varepsilon_k)$, the fact that $\lambda = 1/(2m_f)$, $\bar{d} \leq D$, and $\|z_k - \bar{z}\| \leq D$, and the definition of $\kappa_0$, we conclude that

$$\begin{aligned}\bar{d}\sigma_A^+\|q_k\| &\leq \bar{d}\left(\|\xi_k\| + \nabla_f + 2m_fC_3D\right) \\ &\leq (\bar{d} + \|z_k - \bar{z}\|)L_h + \langle \xi_k, z_k - \bar{z}\rangle + 2m_f\varepsilon_k + \bar{d}\left(\nabla_f + 2m_fC_3D\right) \\ &\leq D\left(2\left[L_h + \nabla_f\right] + \left[4C_3 + C_2\right]m_fD\right) - \langle q_k, Az_k - b\rangle \\ &= D\kappa_0 - \langle q_k, Az_k - b\rangle.\end{aligned}$$

b) Using part a), the definition of $q_k$, and the Cauchy-Schwarz and triangle inequalities, we have that

$$\begin{aligned}c\bar{d}\sigma_A^+\|Az_k - b\| &= \bar{d}\sigma_A^+\|q_k - p_{k-1}\| \leq \bar{d}\sigma_A^+\|q_k\| + \bar{d}\sigma_A^+\|p_{k-1}\| \\ &\leq D\kappa_0 - \langle q_k, Az_k - b\rangle + \bar{d}\sigma_A^+\|p_{k-1}\| \\ &\overset{(2.9)}{=} D\kappa_0 - \langle p_{k-1}, Az_k - b\rangle - c\|Az_k - b\|^2 + \bar{d}\sigma_A^+\|p_{k-1}\| \\ &\leq D\kappa_0 - c\|Az_k - b\|^2 + \|p_{k-1}\|\left(\|Az_k - b\| + \bar{d}\sigma_A^+\right).\end{aligned}$$

Moving the $-c\|Az_k - b\|^2$ term to the left-hand-side, dividing the resulting inequality by $\|Az_k - b\| + \bar{d}\sigma_A^+$, and using the definition of $\theta_D$ in (2.18) we conclude that

$$c\|Az_k - b\| \leq \frac{D\kappa_0}{\|Az_k - b\| + \bar{d}\sigma_A^+} + \|p_{k-1}\| \leq \frac{\theta_D\kappa_0}{\sigma_A^+} + \|p_{k-1}\|.$$

c) Let $k \equiv 1 \bmod k_c$. Using part a) and the fact that $q_k = p_k = p_{k-1} + c(Az_k - b)$, we have that

$$\bar{d}\sigma_A^+\|p_k\| = \bar{d}\sigma_A^+\|q_k\| \leq D\kappa_0 - \langle q_k, Az_k - b\rangle = D\kappa_0 + \frac{1}{c}\langle p_k, p_{k-1}\rangle - \frac{1}{c}\|p_k\|^2$$

which implies the desired bound. ∎

We observe that Lemma 3.11(c) always holds under the weaker assumption that $\bar{z} \in \mathcal{H}$ and $A\bar{z} = b$ but the scalar $\bar{d}$ which appears on it becomes zero when $\bar{z} \in \partial\mathcal{H}$. The following technical result establishes the boundedness of the sequence of Lagrange multipliers $\{p_k\}$ if instead **(B4)** is assumed, and hence $\bar{d} > 0$.

PROPOSITION 3.12. *The sequence $\{p_k\}$ generated by S-IAIPAL satisfies*

$$\|p_k\| \leq \frac{\theta_D \kappa_0}{\sigma_A^+} \tag{3.27}$$

*for every $k \geq 0$, where $\kappa_0$ and $\theta_D$ are as in (2.19) and (2.18), respectively.*

*Proof.* The proof is done by induction on $k$. Since $p_0 = 0$ and $\kappa_0 \geq 0$, (3.27) trivially holds for $k = 0$. Assume now that (3.27) holds with $k = k - 1$ for some $k \geq 1$. If $k \not\equiv 1 \bmod \lceil \alpha c \rceil$, then (2.9) implies $p_k = p_{k-1}$, and (3.27) holds by our induction hypothesis. If $k \equiv 1 \bmod \lceil \alpha c \rceil$, then the induction hypothesis together with Lemma 3.11(c), the definition of $\theta_D$ in (2.18), and the Cauchy-Schwarz inequality, imply that

$$\begin{aligned}\left(\frac{\|p_k\|}{c} + \sigma_A^+ \bar{d}\right)\|p_k\| &\leq \frac{\|p_k\|\|p_{k-1}\|}{c} + D\kappa_0 \leq \frac{\|p_k\| D\kappa_0}{c\sigma_A^+ \bar{d}} + D\kappa_0 \\ &= \left(\frac{\|p_k\|}{c} + \sigma_A^+ \bar{d}\right)\frac{\theta_D \kappa_0}{\sigma_A^+},\end{aligned}$$

and hence that $\|p_k\| \leq \theta_D \kappa_0 / \sigma_A^+$. We have thus proved that (3.27) holds for all $k \geq 0$. ∎

The following result establishes that $\|\hat{w}_k\| = \mathcal{O}(\sqrt{\Delta_k} + \alpha^{-1/2} c^{-1})$ and $\|A\hat{z}_k - b\| = \mathcal{O}(c^{-1})$. Since $\Delta_k = \mathcal{O}(k^{-1})$ in view of (3.20) and (3.27), it follows that $\|\hat{w}_k\|$ can be made arbitrarily small as the penalty parameter $c$ increases.

LEMMA 3.13. *The sequence $\{(\hat{z}_k, \hat{w}_k, z_k)\}$ generated by S-IAIPAL satisfies the following bounds:*

*a) $\|A\hat{z}_k - b\| \leq \kappa_2^2 m_f / (c\|A\|^2)$ for every $k \geq 1$;*

*b) $\min_{2 \leq i \leq k} \|\hat{w}_i\|^2 \leq 2m_f C_1 \Delta_k + 2m_f \kappa_1^2 / (c\lceil c\alpha \rceil \|A\|^2)$ for every $k \geq 1$.*

*where $C_1$ is as in (2.6), $\Delta_k$ is as in (2.12), $\kappa_0$ is as in (2.19), and $\kappa_1$ and $\kappa_2$ are as in (2.20).*

*Proof.* a) It follows from Lemma 3.11(b), the second inequality in (3.3), the triangle inequality, and the definitions of $(\sigma_c, L_c)$ and $p_k$ given in (2.6) and (2.9), respectively, that

$$\begin{aligned}\|A\hat{z}_k - b\| \leq \|Az_k - b\| + \|A\|\|\hat{z}_k - z_k\| &\leq \frac{\theta_D \kappa_0}{\sigma_A^+ c} + \frac{\|p_{k-1}\|}{c} + \frac{\sigma_c \|A\|\|r_k\|}{\sqrt{\lambda L_c + 1}} \\ &\leq \frac{2\theta_D \kappa_0}{\sigma_A^+ c} + \frac{\nu \|A\|\|r_k\|}{\lambda L_c + 1} \leq \frac{2\theta_D \kappa_0}{\sigma_A^+ c} + \frac{\nu \|r_k\|}{\lambda c \|A\|},\end{aligned}$$

where the last inequality is due to $L_c \geq c\|A\|^2$. It follows from the above inequalities, (3.27), and the first inequality in (3.21) that

$$\frac{c\|A\|^2}{m_f}\|A\hat{z}_k - b\| \leq \frac{1}{m_f}\left(\frac{2\|A\|^2 \theta_D \kappa_0}{\sigma_A^+} + \frac{\nu \|A\| D}{\lambda(1-\sigma)}\right) = 2\theta_A \theta_D \left(\frac{\|A\|\kappa_0}{m_f} + \frac{\nu \sigma_A^+ \bar{d}}{1-\sigma}\right) \tag{3.28}$$

where the last relation is due to the definitions of $\lambda$, $\theta_A$, and $\theta_D$ given in (2.6) and (2.18). On the other hand, using the definitions of $C_3$ and $\kappa_0$ given in (2.16) and (2.19), respectively, and the fact that $\bar{d} \leq D$ and $\sigma_A^+ \leq \|A\|$, we have

$$\frac{\nu \sigma_A^+ \bar{d}}{1-\sigma} \leq C_3 \|A\| D \leq \frac{\|A\|\kappa_0}{4m_f}.$$

Hence, the desired bound immediately follows from (3.28), the latter inequalities and the definition of $\kappa_2$ in (2.20).

b) Define $I(k) := \{i : p_i \neq p_{i-1}, 2 \leq i \leq k\}$. In view of the multiplier update rule of S-IAIPAL, it is straightforward to show that $|I(k)| \leq \lfloor (k-1)/\lceil \alpha c \rceil \rfloor \leq 2(k-1)/\lceil \alpha c \rceil$. Using (2.18), (3.16), (3.20), (3.27), the relation $(a+b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \Re$, and the previous bound on $|I(k)|$, we have

$$\begin{aligned}
\frac{\lambda}{C_1} \min_{2\leq i\leq k} \|\hat{w}_i\|^2 &\leq \Delta_k + \sum_{i=2}^{k} \frac{\|\Delta p_i\|^2}{c(k-1)} = \Delta_k + \sum_{i\in I(k)} \frac{\|\Delta p_i\|^2}{c(k-1)} \\
&\leq \Delta_k + \frac{2\sum_{i\in I(k)}(\|p_i\|^2 + \|p_{i-1}\|^2)}{c(k-1)} \leq \Delta_k + \left[\frac{\theta_D \kappa_0}{\sigma_A^+}\right]^2 \left[\frac{4|I(k)|}{c(k-1)}\right] \\
&\leq \Delta_k + \frac{8}{c\lceil c\alpha\rceil}\left[\frac{\theta_D \kappa_0}{\sigma_A^+}\right]^2 = \Delta_k + \frac{8\theta_A^2\theta_D^2\kappa_0^2}{c\lceil c\alpha\rceil \|A\|^2} = \Delta_k + \frac{\kappa_1^2}{c\lceil c\alpha\rceil \|A\|^2 C_1},
\end{aligned}$$

where the last relation is due to the definition of $\kappa_1$ given in (2.20). The desired bound then follows in view of the fact that $\lambda = 1/(2m_f)$. ∎

Notice that, unless $c$ is sufficiently large, the bounds derived in the above lemma does not guarantee that either the feasibility residual $\|A\hat{z}_k - b\|$ or the stationarity residual $\|\hat{w}_k\|$ become sufficiently small, regardless of how large $k$ is. This is in contrast to all of the penalty/AL methods in [18, 19, 20, 24, 26, 30, 40] where the stationarity residual is always sufficiently small whenever the penalty parameter is updated.

We are now ready to present the proof of Theorem 2.3.

*Proof of Theorem 2.3.* a) This statement follows immediately from (3.2).

b) Let $T_0 = T_0(m_f, \hat{\rho})$ where $T_0(\cdot,\cdot)$ is as in (2.21) and assume that S-IAIPAL has reached the $T_0$-th iteration and has not stopped in its step 2. Using (3.20) with $k = T_0$, and the definitions of $\lambda$, $\kappa_1$, and $T_0$ given in (2.6), (2.20), and (2.21), respectively, we then conclude that

$$\begin{aligned}
\Delta_{T_0} &\leq \frac{1}{T_0 - 1}\left[3\left(\Delta\phi^* + 2m_f D^2\right) + \frac{\|p_{T_0}\|^2}{2c}\right] \\
&\leq \frac{1}{T_0 - 1}\left[3\left(\Delta\phi^* + 2m_f D^2\right) + \frac{(\theta_A\theta_D\kappa_0)^2}{2m_f}\right] \\
&= \frac{1}{T_0 - 1}\left[3\left(\Delta\phi^* + 2m_f D^2\right) + \frac{\kappa_1^2}{16C_1 m_f}\right] \leq \frac{\hat{\rho}^2}{4C_1 m_f} = \frac{\lambda\hat{\rho}^2}{2C_1}.
\end{aligned}$$

Hence, S-IAIPAL must stop in step 3 of the $T_0$-th iteration.

c) This statement follows immediately from Lemma 3.13(b) and condition (2.23) on the penalty parameter $c$.

d) First note that it follows from part b) that S-IAIPAL stops either in step 2 or step 3 after a finite number of iteration. Now, in view of the stopping criterion in step 2 and part c), it follows that S-IAIPAL stops with success at the $k$-th iteration if and only if $\hat{w}_k$ satisfies $\|\hat{w}_k\| \leq \hat{\rho}$, in which case the triple $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ is a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1.1) due to part a) and Definition 2.1. Now, assume for contradiction that S-IAIPAL stops in step 3 (instead of step 2) at some iteration $k$, and hence that

$$\min_{2\leq i\leq k} \|\hat{w}_i\| > \hat{\rho}, \qquad \Delta_k \leq \frac{\lambda\hat{\rho}^2}{2C_1} = \frac{\hat{\rho}^2}{4m_f C_1}$$

in view of the last observation above, (2.12), and the definition of $\lambda$ in (2.6). These two inequalities together with (2.23), and Lemma 3.13(c), yield the contradiction

$$\hat{\rho}^2 < \min_{2\le i\le k} \|\hat{w}_i\|^2 \le 2C_1 m_f \Delta_k + \frac{2m_f\kappa_1^2}{c\lceil c\alpha\rceil\|A\|^2} \le \frac{\hat{\rho}^2}{2} + \frac{\hat{\rho}^2}{2} = \hat{\rho}^2.$$

which must mean that S-IAIPAL stops with a $(\hat{\rho},\hat{\eta})$-approximate stationary solution in step 2. ∎

**4. Numerical experiments.** This section presents experiments[7] which benchmark different variants of IAIPAL. The first subsection benchmarks IAIPAL against three other state-of-the-art constrained composite optimization solvers, while the second subsection compares them against the $\mathcal{O}(\varepsilon^{-2})$ complexity method in [42, 43].

We start by describing the details of the IAIPAL variants IPL, IPL(A1), and IPL(A2). All of them use the parameters

$$c_1 = \max\left\{1, \frac{L_f}{\|\mathcal{A}\|^2}\right\}, \quad \sigma = \frac{1}{\sqrt{2}}, \quad \nu = \sqrt{\sigma\,(\lambda L_f + 1)}, \quad \tau = 2.$$

IPL is as described in Subsection 2.3 with $\alpha = 1/\|A\|^2$, while IPL(A1) and IPL(A2) are a modification of IPL where the ACG subroutine is replaced with an adaptive ACG variant whose specific description can be found in [16, Section 5.2]. The difference between the latter ACG variant compared to the first one is that the latter one adapts its proximal gradient step to the local curvature of its objective function (see the discussion in the second paragraph following ACG in Appendix A.1). IPL(A1) chooses $\alpha = 1/\|A\|^2$ while IPL(A2) chooses $\alpha = 1/c$, i.e., a multiplier update is performed at every outer iteration.

We now describe the other methods used in the first subsection, namely, two variants of the QP-AIPP method of [18] (nicknamed QP and QP(A)), a variant of the R-QP-AIPP method of [19] (nicknamed RQP), and the iALM of [24]. QP is the method in [16, Algorithm 4.1.1] while QP(A) is a modification of QP that replaces its ACG subroutine with the same adaptive ACG variant used by IPL(A). RQP is the variant in [16, Algorithm 5.4.1] which adds another level of adaptability to QP(A) in the sense that its prox parameter $\lambda$ is also adapted to the local curvature of the objective function (see the discussion in [19, Section 1]). Our implementation of iALM uses the parameters

$$\sigma = 2, \quad \beta_0 = \max\left\{1, \frac{L_f}{\|\mathcal{A}\|^2}\right\}, \quad w_0 = 1, \quad \boldsymbol{y}^0 = 0, \quad \gamma_k = \frac{(\log 2)\,\|c(x^1)\|}{(k+1)\,[\log(k+2)]^2},$$

for every $k \ge 1$. Moreover, the starting point given to the $k$-th APG call (in the iALM) is set to be $\boldsymbol{x}^{k-1}$, which is the prox center for the $k$-th prox subproblem.

We next describe the other methods used in the second subsection, namely, the three variants of the S-prox-ALM of [42, 43] (nicknamed SPA1–SPA3). The parameter quadruple $(\alpha, p, c, \beta)$ and initial points $(y_0, z_0)$ used by all the three variants are

$$\alpha = \frac{\Gamma}{4}, \quad p = 2(L_f + \Gamma\|A\|^2), \quad c = \frac{1}{2(L_f + \Gamma\|A\|^2)}, \quad \beta = 0.5, \quad y_0 = 0, \quad z_0 = x_0,$$

where $\Gamma = 0.1$, 1, and 10 for SPA1, SPA2, and SPA3, respectively. Note that the choice of $(\alpha, p, c, \beta)$ above with $\Gamma = 10$ is the one that is used in the limited quadratic

[7] See https://github.com/wwkong/nc_opt/tree/master/tests/papers/IAIPAL for the full code.

programming experiments of [43, Section 6.2]. Moreover, the aforementioned reference establishes the iteration-complexity of S-Prox-ALM for a range of sufficiently small parameters $\beta$ that does not necessarily include the assigned value above, i.e, $\beta = 0.5$.

Some additional technical details about the experiments are also as follows. First, all of the tables below report the total number of innermost iterations that each of the methods need to obtain a quadruple satisfying (4.1) below. This is done so that the iteration cost for reported in our experiments are comparable in the sense that each method require $\mathcal{O}(1)$ resolvent and gradient evaluations per iteration. Second, the algorithms are implemented in MATLAB 2021a and are run on a Windows 64-bit machine with two Intel(R) Xeon(R) Gold 6240 processors and 12 GB of RAM. Third, bold text in the tables of this section indicates the method that performed the most efficiently in a particular metric and problem instance. Finally, the log KKT gap in the experiments below refers to the normalized quantity

$$\hat{r} := \log_{10}\left(\max\left\{\frac{\|\hat{w}\|}{1+\|\nabla f(z_0)\|}, \frac{\|A\hat{z}-b\|}{1+\|Az_0-b\|}\right\}\right).$$

**4.1. Quadratic SDP.** This subsection presents the performance of IAIPAL method against several benchmark methods on a set of nonconvex quadratic semidefinite programming (QSDP) problems.

Given a pair of dimensions $(\ell, n) \in \mathbb{N}^2$, a scalar pair $(\omega_1, \omega_2) \in \mathbb{R}^2_{++}$, linear operators $\mathcal{Q} : \mathbb{S}^n_+ \mapsto \mathbb{R}^\ell$, $\mathcal{B} : \mathbb{S}^n_+ \mapsto \mathbb{R}^n$, and $\mathcal{C} : \mathbb{S}^n_+ \mapsto \mathbb{R}^\ell$ defined by

$$[\mathcal{Q}(Z)]_i = \langle Q_i, Z\rangle, \quad [\mathcal{B}(Z)]_j = \langle B_j, Z\rangle, \quad [\mathcal{C}(Z)]_i = \langle C_i, Z\rangle,$$

for matrices $\{Q_i\}_{i=1}^\ell, \{B_j\}_{j=1}^n, \{C_i\}_{i=1}^\ell \subseteq \mathbb{R}^{n\times n}$, positive diagonal matrix $D \in \mathbb{R}^{n\times n}$, and a vector pair $(b, d) \in \mathbb{R}^\ell \times \mathbb{R}^\ell$, this subsection considers the following QSDP:

$$\begin{aligned} \min_Z \ & \left[f(Z) := -\frac{\omega_1}{2}\|D\mathcal{B}(Z)\|^2 + \frac{\omega_2}{2}\|\mathcal{C}(Z)-d\|^2\right] \\ \text{s.t. } & \mathcal{Q}(Z) = b, \quad Z \in P^n, \end{aligned}$$

where $P^n = \{Z \in \mathbb{S}^n_+ : \text{trace}\,(Z) = 1\}$. In particular, the problem instances tested are given in Table 4.1.

We now describe the experiment parameters. First, the dimensions are $(\ell, n) = (30, 100)$ and only 5% of the entries of $Q_i, B_j$, and $C_i$ are nonzero. Second, the entries of $Q_i$, $B_j$, $C_i$, $D$, $b$, and $d$ are generated using the procedure described in [16, Subsection 5.5.2.1]. Fifth, given a starting point $z_0 \in \Re^{n\times n}$, all of the methods attempt to find a quadruple $(\hat{z}, \hat{p}, \hat{w}, \hat{q})$ satisfying $\hat{w} \in \nabla f(\hat{z}) + \partial\delta_{P^n}(\hat{z}) + \mathcal{Q}^*\hat{p}$ and

$$\frac{\|\hat{w}\|}{1+\|\nabla f(z_0)\|} \le \hat{\rho}, \quad \frac{\|\mathcal{Q}\hat{z}-b\|}{1+\|\mathcal{Q}z_0-b\|} \le \hat{\eta}, \tag{4.1}$$

with $\hat{\rho} = \hat{\eta} = 10^{-4}$. Sixth, using the fact that $\|Z\|_F \le 1$ for every $Z \in P_n$, the constant hyperparameters for the IPL and iALM methods are set to $L_g = 0$, $L_j = 0$, $\rho_j = 0$, and $B_j = \|Q_j\|_F$ for $1 \le j \le \ell$. Finally, each problem instance considered is based on a specific pair $(m_f, L_f)$ for which the scalar pair $(\omega_1, \omega_2)$ is selected so that $L_f = \lambda_{\max}(\nabla^2 f)$ and $m_f = -\lambda_{\min}(\nabla^2 f)$.

We now make several observations and conclusions based on these tables. First, comparing the results between QP(A) and RQP, we conclude that the presence of an adaptive prox stepsize search in the latter method considerably improves its performance compared to the former. Second, IPL(A2) is the direct counterpart of QP(A),

| Log KKT Gap / Log Function Value (row 1)<br>Iteration count / Runtime (row 2) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m_f$ | $L_f$ | iALM | QP | QP(A) | RQP | IPL | IPL(A1) | IPL(A2) |
| $10^0$ | $10^2$ | -4.1/-0.98 | -4.2/-0.98 | -4.0/-0.98 | -4.2/-0.98 | -4.0/-0.98 | -4.0/-0.98 | -4.0/-0.98 |
| | | 20.4/10.3 | 4.1/3.8 | 3.2/3.5 | 1.9/2.1 | 3.0/3.1 | **1.4/1.7** | 1.4/1.7 |
| $10^0$ | $10^3$ | -4.1/0.04 | -4.2/0.04 | -4.3/0.04 | -4.2/0.04 | -4.0/0.04 | -4.0/0.04 | -4.0/0.04 |
| | | 36.2/18.6 | 3.9/3.7 | 4.4/4.7 | 2.0/2.2 | 1.7/1.7 | **0.8/0.9** | 0.8/0.9 |
| $10^0$ | $10^4$ | -4.3/1.04 | -4.3/1.04 | -4.3/1.04 | -4.3/1.04 | -4.3/1.04 | -4.3/1.04 | -4.3/1.04 |
| | | 102.1/51.5 | 4.1/3.7 | 9.3/10.1 | 2.5/2.7 | 1.3/1.3 | **0.6/0.8** | 0.6/0.8 |
| $10^1$ | $10^5$ | -4.3/2.04 | -4.3/2.04 | -4.3/2.04 | -4.3/2.04 | -4.3/2.04 | -4.3/2.04 | -4.3/2.04 |
| | | 104.3/52.5 | 4.1/3.8 | 9.3/10.0 | 2.5/2.8 | 3.3/3.4 | 1.5/1.8 | **0.6/0.8** |
| $10^2$ | $10^5$ | -4.3/2.04 | -4.2/2.04 | -4.3/2.04 | -4.2/2.04 | -4.1/2.04 | -4.1/2.04 | -4.0/2.04 |
| | | 48.9/24.7 | 3.9/3.6 | 4.4/4.8 | 2.0/2.2 | 2.4/2.4 | 1.0/1.3 | **0.8/0.9** |
| $10^3$ | $10^5$ | -4.2/2.02 | -4.1/2.02 | -4.0/2.02 | -4.2/2.02 | -4.0/2.02 | -4.0/2.02 | -4.0/2.02 |
| | | 39.8/20.3 | 4.4/4.1 | 3.7/3.9 | 2.1/2.4 | 2.0/2.1 | **0.9/1.1** | 0.9/1.1 |

TABLE 4.1

*Results for constant $m_f$ and variable $L_f$. Within each $(m_f, L_f)$ multirow, the first row presents $\log_{10}$ KKT gaps / $\log_{10}$ function values, while the second row presents iteration counts (in thousands) / runtimes (in tens of seconds). Log function value is computed as $\log_{10}\phi(\hat{z})$.*

but its performance is better than the improved version of QP(A), namely RQP, in nine out of ten problem instances. Third, in view of the first remark above, it is reasonable to infer that IPL(A1) and IPL(A2) could be considerably improved if the prox parameter $\lambda$ is adaptively chosen. As the analysis for such an IPL variant involves several technical difficulties, we leave its development for a future work.

**4.2. Comparison with an $\mathcal{O}(\varepsilon^{-2})$ complexity method.** We start by comparing and contrasting the theoretical properties of each method. First, both the IAIPAL method and the S-prox-ALM are augmented Lagrangian-based methods applied to NCO problems. More specifically, SPA considers (1.1) under the requirement that $h$ is the indicator function of a polyhedron. Second, the S-prox-ALM also considers a sequence of proximal subproblems as in (1.4), and applies a single composite gradient step to inexact solve (1.4) instead of an ACG-type subroutine. Finally, while the IAIPAL method only requires choosing its parameters based on the scalars $m_f$, $L_f$, and $\|A\|$ to guarantee convergence, the S-prox-ALM requires choosing its parameters based on the supremum of a set of Hoffman constants (see the proof of [43, Lemma 3.10] and [43, Lemma 4.8]) that is generally difficult to compute.

We now present some numerical results that compare the S-prox-ALM variants against IP(A1), QP(A), and RQP. Since the S-prox-ALM does not have convergence guarantees for the QSDP problem in Subsection 4.1 (because the domain of $h$ is not polyhedral), we consider the vector variant of the QSDP. More specifically, given a pair of dimensions $(\ell, n) \in \mathbb{N}^2$, a scalar pair $(\omega_1, \omega_2) \in \mathbb{R}^2_{++}$, matrices $Q, C \in \mathbb{R}^{\ell\times n}$ and $B \in \Re^{n\times n}$, positive diagonal matrix $D \in \mathbb{R}^{n\times n}$, and a vector pair $(b, d) \in \mathbb{R}^\ell \times \mathbb{R}^\ell$, we consider the problem

$$\min_z \ \left[ f(z) - \frac{\omega_1}{2}\|DBz\|^2 + \frac{\omega_2}{2}\|\mathcal{C}z - d\|^2 \right]$$

$$\text{s.t. } Qz = b, \quad z \in \Delta^n,$$

where $\Delta^n := \{x \in \Re^n : \sum_{i=1}^n x_i = 1\}$. In particular, the problem instances tested are given in Table 4.2.

| **Log KKT Gap / Log Function Value (row 1)**<br>**Iteration count (row 2)** | | | | | | | |
|---|---|---|---|---|---|---|---|
| $m_f$ | $L_f$ | QP(A) | RQP | IPL(A1) | SPA1 | SPA2 | SPA3 |
| $10^0$ | $10^2$ | -2.0/2.30 | -2.9/2.30 | **-3.2**/2.30 | -2.3/2.30 | -1.7/2.30 | -1.1/2.30 |
| | | 11.5 | 12.2 | 11.6 | 9.9 | 9.8 | 9.9 |
| $10^0$ | $10^3$ | -1.2/2.30 | -2.6/2.30 | **-3.2**/2.30 | -1.9/2.30 | -1.7/2.30 | -1.1/2.30 |
| | | 11.7 | 12.2 | 11.2 | 8.8 | 9.8 | 9.8 |
| $10^0$ | $10^4$ | -0.7/2.30 | -2.3/2.30 | **-4.1**/2.30 | -1.7/2.30 | -1.7/2.30 | -1.5/2.30 |
| | | 11.9 | 12.1 | 13.6 | 9.9 | 9.9 | 9.9 |
| $10^1$ | $10^5$ | -0.7/2.39 | -2.3/2.40 | **-4.1**/2.41 | -0.7/2.39 | -1.4/2.41 | -1.7/2.41 |
| | | 11.9 | 12.2 | 12.2 | 9.9 | 9.9 | 9.8 |
| $10^2$ | $10^5$ | -0.9/2.35 | -2.5/2.37 | **-3.4**/2.37 | -0.7/2.35 | -1.4/2.37 | -1.8/2.37 |
| | | 10.2 | 12.2 | 8.5 | 9.9 | 9.9 | 9.9 |
| $10^3$ | $10^5$ | -1.5/1.81 | -2.4/1.83 | **-7.0**/1.83 | -0.6/1.75 | -1.3/1.87 | -2.0/1.87 |
| | | 9.5 | 12.0 | 3.0 | 9.9 | 9.9 | 9.7 |

TABLE 4.2

*Results for constant $m_f$ and variable $L_f$. Within each $(m_f, L_f)$ multirow, the first row presents $\log_{10}$ KKT gaps / $\log_{10}$ function values, while the second row presents iteration counts (in thousands). All experiments were run until 600 seconds have passed. Log function value is computed as $\log_{10}(200 + \phi(\hat{z}))$.*

We now describe the experiment parameters for the problem instances considered. First, the dimension pair is $(\ell, n) = (20, 1000)$ and all generated matrices have full density. Second, the entries of $Q$, $B$, $C$, and $d$ (resp. $D$) are generated by sampling from the uniform distribution $\mathcal{U}[0, 1]$ (resp. $\mathcal{U}\{1, ..., 1000\}$). Third, the vector $b$ is set to $b = Q(\boldsymbol{e}/n)$ where $\boldsymbol{e}$ is a vector of all ones. Fourth, the initial starting point $z_0$ is a set to be $\tilde{z}/\sum_{i=1}^n \tilde{z}_i$ where the entries of $\tilde{z}$ are sampled from the $\mathcal{U}[0, 1]$ distribution. Fifth, given a starting point $z_0 \in \Re^n$, all of the methods attempt to find a quadruple $(\hat{z}, \hat{p}, \hat{w}, \hat{q})$ satisfying $\hat{w} \in \nabla f(\hat{z}) + \partial \delta_{\Delta^n}(\hat{z}) + Q^* \hat{p}$ and (4.1) with $\hat{\rho} = \hat{\eta} = 10^{-7}$. Finally, all experiments are run with a time limit of 600 seconds.

From the results, we can see that the IPL(A2) variant is substantially more efficient than SPA1–SPA3, QP(A), and RQP. We also notice that SPA1 (resp. SPA3) tends to perform better when $L_f$ is small (large).

**5. Concluding remarks.** This paper proposes the IAIPAL method for finding a $(\hat{\rho}, \hat{\eta})$-approximate stationary point (see Definition 2.1) of a class of linearly-constrained smooth NCO problems, and establishes, up to logarithmic terms, an $\mathcal{O}(\hat{\rho}^{-5/2} + \hat{\rho}^{-2}\hat{\eta}^{-1/2})$ ACG iteration complexity bound for it. Moreover, IAIPAL is the first PAL method with provable complexity bounds for the case where $(\theta, \chi_k) = (0, 1)$ in (1.6) for every $k \geq 1$. Computational results also show that IAIPAL substantially outperforms other algorithms in the literature for solving (1.1) (or special cases of it).

We now discuss some possible extensions of our paper. First, it is worth developing an adaptive variant of IAIPAL as described in the conclusion of Section 4. Second,

one could analyze the convergence and computational behavior of IAIPAL under the multiplier update rule $p_{k+1} = p_k + \chi c(Az_k - b)$ where $\chi$ is a positive scalar lying in a certain range. Finally, it is worth investigating whether the iteration-complexity of IAIPAL can be improved, possibly for special instances of (1.1).

**Appendix A. Other technical results.** This section is divided into three subsections. The first one revise an accelerated gradient method used for solving the IAIPAL subproblems. The second subsection establishes a result, using convex analysis, that is used to prove Lemma 3.10. The last subsection presents a result regarding a refinement procedure related to the pair $(\hat{z}, \hat{w})$ computed in step 2 of S-IAIPAL.

**A.1. An accelerated composite gradient method.** Consider the following composite optimization problem

$$\min\{\psi(x) := \psi_s(x) + \psi_n(x) : x \in \Re^n\} \tag{A.1}$$

where the following conditions are assumed to hold:

**(A1)** $\psi_n : \Re^n \to (-\infty, +\infty]$ is a proper closed convex function;

**(A2)** $\psi_s$ is a convex differentiable function on $\operatorname{dom}\psi_n$ and there exists $(\widetilde{\mu}, \widetilde{M}) \in \Re^2_+$ satisfying $\widetilde{M} > \widetilde{\mu}$ and $\widetilde{\mu}\|u-x\|^2/2 \le \psi_s(u) - \ell_{\psi_s}(u;x) \le \widetilde{M}\|u-x\|^2/2$ for every $x, u \in \operatorname{dom}\psi_n$, where $\ell_{\psi_s}(\cdot\,;\cdot)$ is defined in (1.9).

We are now ready to state ACG. It is worth mentioning that other ACG variants such as the ones in [1, 11, 34, 33] could also be used in the development of IAIPAL.

---

**ACG**

---

(0) Let a pair of functions $(\psi_s, \psi_n)$ satisfying **(A1)** and **(A2)** for some $(\widetilde{\mu}, \widetilde{M}) \in \Re^2_+$, a scalar $\tilde{\sigma} > 0$, and an initial point $y_0 \in \operatorname{dom}\psi_n$ be given; set $x_0 = y_0$, $A_0 = 0$, $\tau_0 = 1$, $\zeta = 1/(\widetilde{M} - \widetilde{\mu})$, and $j = 0$;

(1) compute the iterates

$$a_j = \frac{\zeta\tau_j + \sqrt{(\zeta\tau_j)^2 + 4\tau_j A_j}}{2}, \quad A_{j+1} = A_j + a_j, \quad \tilde{x}_j = \frac{A_j y_j + a_j x_j}{A_{j+1}}$$

$$\tau_{j+1} = \tau_j + \widetilde{\mu} a_j, \quad y_{j+1} = \operatorname*{argmin}_{y\in\Re^n}\left\{\ell_{\psi_s}(y;\tilde{x}_j) + \psi_n(y) + \frac{\widetilde{M}}{2}\|y - \tilde{x}_j\|^2\right\},$$

$$x_{j+1} = \frac{1}{\tau_{j+1}}\left[\frac{a_j}{\zeta}(y_{j+1} - \tilde{x}_j) + \widetilde{\mu} a_j y_{j+1} + \tau_j x_j\right];$$

(2) compute the quantities

$$u_{j+1} = \widetilde{\mu}(y_{j+1} - x_{j+1}) + \frac{x_0 - x_{j+1}}{A_{j+1}},$$

$$\eta_{j+1} = \frac{1}{2A_{j+1}}\left(\|x_0 - y_{j+1}\|^2 - \tau_{j+1}\|x_{j+1} - y_{j+1}\|^2\right);$$

(3) if the inequality

$$\|u_{j+1}\|^2 + 2\eta_{j+1} \le \tilde{\sigma}^2\|y_0 - y_{j+1} + u_{j+1}\|^2$$

holds, then stop and output $(y, u, \eta) := (y_{j+1}, u_{j+1}, \eta_{j+1})$; otherwise, set $j = j + 1$ and go to (1).

Some remarks about ACG follow. First, the most common way of describing an iteration of ACG is as in step 1. Second, the auxiliary iterates $\{u_j\}$ and $\{\eta_j\}$ computed in step 2 are used to develop a stopping criterion for ACG when it is called as a subroutine for solving the subproblems generated in step 1 of S-IAIPAL in Subsection 2.2. Third, it can be shown (see for example [7, 17]) that ACG (without steps 2 and 3) with $\widetilde{\mu} = 0$ corresponds to the well-known FISTA algorithm. Fourth, the sequence $\{A_j\}$ has the following increasing property:

$$A_j \geq \frac{1}{\widetilde{M} - \widetilde{\mu}} \max\left\{\frac{j^2}{4}, \left(1 + \sqrt{\frac{\widetilde{\mu}}{4(\widetilde{M} - \widetilde{\mu})}}\right)^{2(j-1)}\right\}, \qquad \forall j \geq 1.$$

Finally, it is worth mentioning that adaptive variants[8] of ACG have been studied, for example, in [4, 16, 27, 34, 35]. A simple level of adaptiveness used in these variants, which is also used inside some of the methods benchmarked in Section 4, is to replace $\widetilde{M}$ in the computation of $y_j$ in step 1 by an estimate $M_j$ computed as follows: $M_j$ is initially set to be $M_{j-1}$ and, if necessary, is repeatedly increased (either additively, multiplicatively, or both) until the inequality $\psi_s(y_j) - \ell_{\psi_s}(y_j; \tilde{x}_{j-1}) \leq M_j\|y_j - \tilde{x}_{j-1}\|^2/2$ is satisfied.

The next result, whose proof can be found in [17, Lemma 2.13], summarizes the main properties of ACG used in this paper.

PROPOSITION A.1. *Let $\{(y_j, u_j, \eta_j)\}_{j\geq 1}$ be the sequence generated by ACG applied to (A.1), where $(\psi_s, \psi_n)$ is a given pair of data functions satisfying* **(A1)** *and* **(A2)**. *Then, the following statements hold:*

*a) for every $j \geq 1$, we have $u_j \in \partial_{\eta_j}(\psi_s + \psi_n)(y_j)$;*

*b) for any $\tilde{\sigma} > 0$, the ACG method outputs a triple $(y, u, \eta)$ satisfying*

$$u \in \partial_\eta(\psi_s + \psi_n)(y) \quad \|u\|^2 + 2\eta \leq \tilde{\sigma}^2\|y_0 - y + u\|^2$$

*in at most*

$$\left\lceil \left(\frac{1}{2} + \sqrt{\frac{\widetilde{M} - \widetilde{\mu}}{\widetilde{\mu}}}\right) \log_1^+\left(\left[\widetilde{M} - \widetilde{\mu}\right]\mathcal{A}_{\widetilde{\mu},\tilde{\sigma}}\right) + 1 \right\rceil \tag{A.2}$$

*iterations, where $\mathcal{A}_{\widetilde{\mu},\tilde{\sigma}} := (2\widetilde{\mu} + 3)(1 + \tilde{\sigma})^2/\tilde{\sigma}^2$.*

**A.2. A convex analysis result.** This subsection contains a technical result of convex analysis. It derives several characterizations of condition **(B2)** and establishes an important inclusion that is used in the proof of Lemma 3.10.

LEMMA A.2. *Let $h \in \overline{\text{Conv}}\,(\Re^n)$ and $L_h \geq 0$ be given. Then, the following statements are equivalent:*

*a) for every $z, z' \in \mathcal{H}$, we have $h(z') \leq h(z) + L_h\|z' - z\|$;*

*b) for every $z, z' \in \mathcal{H}$, we have $h'(z; z' - z) \leq L_h\|z' - z\|$;*

*c) for every $z, z' \in \mathcal{H}$ and $s \in \partial h(z)$, we have $\langle s, z' - z\rangle \leq L_h\|z' - z\|$;*

*d) for every $z \in \mathcal{H}$, we have $\partial h(z) \subset \bar{B}(0; L_h) + N_{\mathcal{H}}(z)$;*

*e) for every $z \in \mathcal{H}$, we have $\partial h(z) \cap \bar{B}(0; L_h) \neq \emptyset$.*

*Moreover, any of the above conditions imply that:*

[8] The closest variant to ACG in this paper can be found in [16, Section 5.2].

*i) $\mathcal{H}$ is closed;*

*ii) for any $z \in \mathcal{H}$ and $\varepsilon \geq 0$, we have $\partial_\varepsilon h(z) \subset \bar{B}(0; L_h) + N^\varepsilon_{\mathcal{H}}(z)$.*

*Proof.* $[a) \Rightarrow b)]$ This statement follows from the fact that $h(z') - h(z) \geq h'(z; z' - z)$ for every $z, z' \in \mathcal{H}$ (see [37, Theorem 23.1]).
$[b) \Rightarrow c)]$ This statement follows from the fact that $h'(z; z' - z) \geq \langle s, z' - z\rangle$ for every $z, z' \in \mathcal{H}$ and $s \in \partial h(z)$, (see [37, Theorem 23.2]).
$[c) \Rightarrow d)]$ Letting $T_{\mathcal{H}}(z) = \mathrm{cl}\,(\mathbb{R}_+ \cdot (\mathcal{H} - z))$ and $N_{\mathcal{H}}(z)$ denote the tangent cone and normal cone of $\mathcal{H}$ at $z$, respectively, and letting $S := \bar{B}(0; L_h) + N_{\mathcal{H}}(z)$, we easily see that c) is equivalent to

$$\langle s, \cdot\rangle \leq L_h\|\cdot\| + I_{T_{\mathcal{H}}(z)}(\cdot) = \sigma_{\bar{B}(0;L_h)}(\cdot) + \sigma_{N_{\mathcal{H}}(z)}(\cdot) = \sigma_S(\cdot) \quad \forall s \in \partial h(z),$$

where the first equality follows in view of the discussion in page 115 of [37] and [12, Example 2.3.1 combined with Proposition 5.2.4], the last equality is due to [37, Corollary 16.4.1]. Since the above hold for every $s \in \partial h(z)$, we conclude that $\sigma_{\partial h(z)} \leq \sigma_S$. Since both $\partial h(z)$ and $S$ are closed, it follows from [37, Corollary 13.1.1] that $\partial h(z) \subset S = \bar{B}(0; L_h) + N_{\mathcal{H}}(z)$.
$[d) \Rightarrow e)]$ Assume that d) holds. We will first show that e) holds for every $z \in \mathrm{ri}\,\mathcal{H}$. Indeed, assume that $z \in \mathrm{ri}\,\mathcal{H}$. This implies that $N_{\mathcal{H}}(z)$ is a subspace, namely, the one orthogonal to the subspace parallel to the affine hull of $\mathcal{H}$. It follows from d) that there exists $s \in \partial h(z)$ and $n \in N_{\mathcal{H}}(z)$ such that $\|s - n\| \leq L_h$. Since $N_{\mathcal{H}}(z)$ is a subspace, it follows that $-n \in N_{\mathcal{H}}(z)$. The claim now follows by the observation that $s \in \partial f(z)$ and $-n \in N_{\mathcal{H}}(z)$ immediately implies that $s - n \in \partial f(z)$. We will now show that e) also holds for every $z \in \mathrm{rbd}\,\mathcal{H}$. Indeed, assume that $z \in \mathrm{rbd}\,\mathcal{H}$. Then, due to [12, Proposition 2.1.8], there exists $\{z_k\} \subset \mathrm{ri}\,\mathcal{H}$ such that $z_k$ converges to $z$ as $k \to \infty$. Since e) holds for every $z \in \mathrm{ri}\,\mathcal{H}$ and $\{z_k\} \subset \mathrm{ri}\,\mathcal{H}$, we conclude that for every $k$, there exists $s_k \in \partial h(z_k)$ such that $\|s_k\| \leq L_h$. Hence, by Bolzano-Weisstrass' theorem, there exists a subsequence $\{s_k\}_{k\in\mathcal{K}}$ converging to some $s$, which clearly satisfies $\|s\| \leq L_h$. Using the fact that $\{(z_k, s_k)\}_{k\in\mathcal{K}} \in \mathrm{Gr}\,(\partial h)$ and $\{(z_k, s_k)\}_{k\in\mathcal{K}}$ converges to $(z, s)$, and the fact that $h \in \overline{\mathrm{Conv}}\,(\Re^n)$ implies that the set $\mathrm{Gr}\,(\partial h)$ is closed, we then conclude that $(z, s) \in \mathrm{Gr}\,(\partial h)$, i.e., $s \in \partial h(z)$. We have thus shown that e) holds for every $z \in \mathrm{rbd}\,\mathcal{H}$ as well.
$[e) \Rightarrow a)]$ Let $z, z' \in \mathcal{H}$ be given and assume that e) holds. Then, there exists $s' \in \partial h(z')$ such that $\|s'\| \leq L_h$. Hence, $h(z) - h(z') \geq \langle s', z - z'\rangle \geq -\|s'\|\,\|z' - z\| \geq -L_h\|z' - z\|$, which proves a).
$[a) \Rightarrow i)]$ Assume that $\{z_k\} \subset \mathcal{H}$ converges to $z$. The fact that $h \in \overline{\mathrm{Conv}}\,(\Re^n)$ and the assumption that (a) holds imply that

$$h(z) \leq \liminf_{k\to+\infty} h(z_k) \leq \liminf_{k\to+\infty} (h(z_1) + L_h\|z_k - z_1\|) = h(z_1) + L_h\|z - z_1\| < +\infty,$$

and hence that $z \in \mathcal{H}$. We have thus shown that $\mathcal{H}$ is closed.
$[a) \Rightarrow ii)]$ Let $z \in \mathcal{H}$ and $\varepsilon \geq 0$ be given and assume that a) holds. Consider the function $\phi_z$ defined as

$$\phi_z(z') := h(z) + L_h\|z' - z\| + I_{\mathcal{H}}(z') \quad \forall z' \in \Re^n.$$

Clearly, $\phi_z(z) = h(z)$ and $\phi_z \geq h$ in view of a). Using these two observations and the definition of the $\varepsilon$-subdifferential given in (1.8), we immediately see that $\partial_\varepsilon h(z) \subset \partial_\varepsilon \phi_z(z)$. On the other hand, using the $\varepsilon$-subdifferential rule for the sum of two convex functions (see [13, Theorem 3.1.1]), we have that

$$\partial_\varepsilon \phi_z(z) \subset \partial_\varepsilon \left(L_h\|\cdot - z\|\right)(z) + \partial_\varepsilon I_{\mathcal{H}}(z) = \partial_\varepsilon \left(L_h\|\cdot\|\right)(0) + N^\varepsilon_{\mathcal{H}}(z),$$

where the last equality is due to the the affine composition rule for the $\varepsilon$-subdifferential (see [13, Theorem 3.2.1]) and the fact that $N^{\varepsilon}_{\mathcal{H}}(\cdot) = \partial_{\varepsilon} I_{\mathcal{H}}(\cdot)$. The implication now follows from the above two inclusions and the fact that $\partial_{\varepsilon}\left(L_h\|\cdot\|\right)(0) = \bar{B}(0; L_h)$. ∎

We observe that a) of Lemma A.2 is the same as condition **(B2)**. Conditions b) to e) are all equivalent to a), and hence **(B2)**. The implication a) ⇒ ii) is the one that is used in the proof of Lemma 3.10.

**A.3. A basic refinement result.** Even though the result below, which is used to prove Proposition 3.1, is a slight variant of [10, Lemma 32], we include its proof for the sake of completeness.

LEMMA A.3. *Assume that* $\tilde{h} \in \overline{\mathrm{Conv}}\,(\Re^n)$, $\tilde{g}$ *is a differentiable function on* $\mathrm{dom}\,\tilde{h}$, *and* $(z, \varepsilon) \in \mathrm{dom}\,\tilde{h} \times \Re_+$ *is such that*

$$0 \in \partial_{\varepsilon}(\tilde{g} + \tilde{h})(z). \tag{A.3}$$

*Assume also that there exists* $\tilde{L} > 0$ *such that*

$$\tilde{g}(u) - \ell_{\tilde{g}}(u; z) \leq \frac{\tilde{L}}{2}\|u - z\|^2 \qquad \forall u \in \mathrm{dom}\,\tilde{h}, \tag{A.4}$$

*and define*

$$\tilde{z} := \operatorname*{argmin}_{u} \left\{ \ell_{\tilde{g}}(u; z) + \tilde{h}(u) + \frac{\tilde{L}}{2}\|u - z\|^2 \right\}, \qquad \widetilde{w} := \tilde{L}(z - \tilde{z}). \tag{A.5}$$

*Then, the quadruple* $(z, \tilde{z}, \widetilde{w}, \varepsilon)$ *satisfies*

$$\widetilde{w} \in \nabla\tilde{g}(z) + \partial\tilde{h}(\tilde{z}), \qquad \widetilde{w} \in \nabla\tilde{g}(z) + \partial_{\varepsilon}\tilde{h}(z), \qquad \|\widetilde{w}\| \leq \sqrt{2\tilde{L}\varepsilon}. \tag{A.6}$$

*Proof.* The first inclusion in (A.6) follows from the definition of $\widetilde{w}$ and the optimality condition for the problem in (A.5). Now, using the first inclusion in (A.6), the definition of $\widetilde{w}$ in (A.5), inclusion (A.3), inequality (A.4), and the subdifferential definition (1.8), we conclude that for every $u \in \Re^n$,

$$\begin{aligned}
h(u) &\geq h(\tilde{z}) + \langle \widetilde{w} - \nabla\tilde{g}(z), u - \tilde{z}\rangle \\
&= h(z) + \langle \widetilde{w} - \nabla\tilde{g}(z), u - z\rangle + h(\tilde{z}) - h(z) + \frac{\|\widetilde{w}\|^2}{\tilde{L}} + \langle \nabla\tilde{g}(z), \tilde{z} - z\rangle \\
&\geq h(z) + \langle \widetilde{w} - \nabla\tilde{g}(z), u - z\rangle + h(\tilde{z}) - h(z) + \frac{\|\widetilde{w}\|^2}{\tilde{L}} + g(\tilde{z}) - g(z) - \frac{\tilde{L}}{2}\|\tilde{z} - z\|^2 \\
&\geq h(z) + \langle \widetilde{w} - \nabla\tilde{g}(z), u - z\rangle - \varepsilon + \frac{\|\widetilde{w}\|^2}{2\tilde{L}}
\end{aligned}$$

which, in view of (1.8), clearly implies the second inclusion in (A.6). Finally, the inequality in (A.6) follows from the above relations with $u = z$. ∎

## REFERENCES


[1] H. Attouch and J. Peypouquet. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. *SIAM J. Optim.*, 26(3):1824–1834, 2016.

[2] N.S. Aybat and G. Iyengar. A first-order smoothed penalty method for compressed sensing. *SIAM J. Optim.*, 21(1):287–313, 2011.

[3] N.S. Aybat and G. Iyengar. A first-order augmented Lagrangian method for compressed sensing. *SIAM J. Optim.*, 22(2):429–459, 2012.

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[5] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York, 1982.

[6] D. Boob, Q. Deng, and G. Lan. Stochastic first-order methods for convex and nonconvex functional constrained optimization. *arXiv:1908.02734*, 2019.

[7] M. I. Florea and S. A. Vorobyov. An accelerated composite gradient method for large-scale composite objective problems. *IEEE Transactions on Signal Processing*, 67(2):444–459, 2018.

[8] M.L.N. Gonçalves, J.G. Melo, and R.D.C. Monteiro. Convergence rate bounds for a proximal admm with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *Pac. J. Optim.*, 15(3):379–398, 2019.

[9] D. Hajinezhad and M. Hong. Perturbed proximal primal–dual algorithm for nonconvex nonsmooth optimization. *Math. Program.*, 176:207–245, 2019.

[10] Y. He and R.D.C. Monteiro. Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player Nash equilibrium problems. *SIAM J. Optim.*, 25(4):2182–2211, 2015.

[11] Y. He and R.D.C. Monteiro. An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM J. Optim.*, 26(1):29–56, 2016.

[12] J.B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I*. Springer, Berlin, 1993.

[13] J.B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms II*. Springer, Berlin, 1993.

[14] M. Hong. Decomposing linearly constrained nonconvex problems by a proximal primal dual approach: algorithms, convergence, and applications. *arXiv:1604.00543*, 2016.

[15] B. Jiang, T. Lin, S. Ma, and S. Zhang. Structured nonconvex and nonsmooth optimization algorithms and iteration complexity analysis. *Comput. Optim. Appl.*, 72(3):115–157, 2019.

[16] W. Kong. Accelerated inexact first-order methods for solving nonconvex composite optimization problems. *arXiv:2104.09685*, April 2021.

[17] W. Kong, J. G. Melo, and R.D.C. Monteiro. FISTA and Extensions - Review and New Insights. *Optimization Online*, 2021.

[18] W. Kong, J.G. Melo, and R.D.C. Monteiro. Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs. *SIAM J. Optim.*, 29(4):2566–2593, 2019.

[19] W. Kong, J.G. Melo, and R.D.C. Monteiro. An efficient adaptive accelerated inexact proximal point method for solving linearly constrained nonconvex composite problems. *Comput. Optim. Appl.*, 76(2):305–346, 2019.

[20] W. Kong and R.D.C. Monteiro. An accelerated inexact proximal point method for solving nonconvex-concave min-max problems. *arXiv:1905.13433v2*, 2019.

[21] G. Lan and R.D.C. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. *Math. Program.*, 138(1):115–139, Apr 2013.

[22] G. Lan and R.D.C. Monteiro. Iteration-complexity of first-order augmented Lagrangian methods for convex programming. *Math. Program.*, 155(1):511–547, Jan 2016.

[23] F. Li and Z. Qu. An inexact proximal augmented Lagrangian framework with arbitrary linearly convergent inner solver for composite convex optimization. *arXiv:1909.09582*, 2019.

[24] Z. Li, P.-Y. Chen, S. Liu, S. Lu, and Y. Xu. Rate-improved inexact augmented Lagrangian method for constrained nonconvex optimization. *Proc. 24th Int. Conf. Artif. Intell. and Statist.*, 130:170–2178, 2021.

[25] Z. Li and Y. Xu. First-order inexact augmented Lagrangian methods for convex and nonconvex programs: nonergodic convergence and iteration complexity. *Preprint*, 2019.

[26] Q. Lin, R. Ma, and Y. Xu. Inexact proximal-point penalty methods for non-convex optimization with non-convex constraints. *arXiv:1908.11518v4*, 2020.

[27] Q. Lin and L. Xiao. An adaptive accelerated proximal gradient method and its homotopy continuation for sparse optimization. *Proc. 31st Int. Conf. Mach. Learn.*, 32:73–81, 2014.

[28] Y.F. Liu, X. Liu, and S. Ma. On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. *Math. Oper. Res.*, 44(2):632–650, 2019.

[29] Z. Lu and Z. Zhou. Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. *arXiv:1803.09941*, 2018.

[30] J.G. Melo, R.D.C. Monteiro, and H. Wang. Iteration-complexity of an inexact proximal accel-

erated augmented Lagrangian method for solving linearly constrained smooth nonconvex composite optimization problems. *arXiv:2006.08048*, 2020.

[31] R.D.C. Monteiro, Ortiz, and Benar F. Svaiter. An adaptive accelerated first-order method for convex optimization. *Comput. Optim. Appl.*, 64:31–73, 2016.

[32] I. Necoara, A. Patrascu, and F. Glineur. Complexity of first-order inexact Lagrangian and penalty methods for conic convex programming. *Optim. Methods Softw.*, pages 1–31, 2017.

[33] Y. E. Nesterov. *Introductory lectures on convex optimization : a basic course*. Kluwer Academic Publ., 2004.

[34] Y.E. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140:1–37, 2013.

[35] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

[36] A. Patrascu, I. Necoara, and Q. Tran-Dinh. Adaptive inexact fast augmented Lagrangian methods for constrained convex optimization. *Optim. Lett.*, 11(3):609–626, 2017.

[37] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[38] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

[39] R.T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research*, 1(2):97–116, 1976.

[40] M. Sahin, A. Eftekhari, A. Alacaoglu, F. Latorre, and V Cevher. An inexact augmented Lagrangian framework for nonconvex optimization with nonlinear constraints. *Adv. Neural Inf. Process. Syst.*, pages 13943–13955, 2019.

[41] Y. Xu. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. *Math. Program.*, 185:199–244, 2021.

[42] J. Zhang and Z.-Q. Luo. A global dual error bound and its application to the analysis of linearly constrained nonconvex optimization. *arXiv:2006.16440*, 2020.

[43] J. Zhang and Z.-Q. Luo. A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization. *SIAM J. Optim.*, 30(3):2272–2302, 2020.